RECOMBINEERING IN MYCOBACTERIA USING MYCOBACTERIOPHAGE PROTEINS

by

Julia Catherine van Kessel

B.S. Biology, Utica College of Syracuse University, 2003

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Julia Catherine van Kessel

It was defended on

July 24, 2008

and approved by

Roger W. Hendrix, Ph.D., Biological Sciences, University of Pittsburgh

William R. Jacobs, Jr., Ph.D., Albert Einstein College of Medicine

Jeffrey G. Lawrence, Ph.D., Biological Sciences, University of Pittsburgh

Valerie Oke, Ph.D., Biological Sciences, University of Pittsburgh

Dissertation Advisor: Graham F. Hatfull, Ph.D., Biological Sciences, University of Pittsburgh

RECOMBINEERING IN MYCOBACTERIA USING

MYCOBACTERIOPHAGE PROTEINS

Julia Catherine van Kessel, Ph.D.

University of Pittsburgh, 2008

Genetic manipulations of *Mycobacterium tuberculosis* are complicated by its slow growth, inefficient DNA uptake, and relatively high levels of illegitimate recombination. Most methods for construction of gene replacement mutants are lengthy and complicated, and the lack of generalized transducing phages that infect *M. tuberculosis* prevents simple construction of isogenic mutant strains. Characterization and genomic analysis of mycobacteriophages has provided numerous molecular and genetic tools for the mycobacteria. Recently, genes encoding homologues of the *Escherichia coli* Rac prophage RecET proteins were revealed in the genome of mycobacteriophage Chec9c. RecE and RecT are functional analogues of the phage λ Red recombination proteins, Exo (exonuclease) and Beta (recombinase), respectively. These recombination enzymes act coordinately to promote high levels of recombination *in vivo* in *E. coli* and related bacteria using short regions of homology, facilitating the development of a powerful genetic technique called 'recombineering.'

Biochemical characterization of Che9c gp60 and gp61 demonstrated that they possess exonuclease and DNA binding activities, respectively, similar to RecET and λ Exo/Beta. Expression of gp60/gp61 in *M. smegmatis* and *M. tuberculosis* substantially increases homologous recombination such that 90% of recovered colonies are the desired gene replacement mutants. Further development of this system demonstrated that Che9c gp61

facilitates introduction of selectable and non-selectable point mutations on mycobacterial genomes at high frequencies using short (<50 nt) ssDNA substrates.

The mycobacterial recombineering system provides a simple and efficient method for mutagenesis with minimal DNA manipulation. While it is clear that similar phage-encoded recombinase homologues are rare, they can be readily identified by genomic studies and by *in vivo* characterization. Several putative recombination systems have been identified in mycobacteriophages Halo, BPs, and Giles, and recombineering of drug-resistance point mutations provides an easy assay for recombinase activity. Analysis of recombinases from various phages – including λ Beta and *E. coli* RecT – indicates that these proteins function best in their native bacteria. The mycobacteriophage-encoded proteins exhibited varying levels of activity, suggesting that analysis of multiple proteins is required to achieve optimal recombination frequencies. The apparent species-specific nature of these recombinases suggests the recombineering technology could likely be extended to any bacterial system through characterization of host-specific bacteriophages.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

The contents of this dissertation, with some additions and alterations, were published previously in references [227-229]. They are reprinted with permission following the guidelines of (1) the Nature Publishing Group with license number 1935451191994, (2) the Journals Rights and Permissions Controller of Blackwell Publishing, Ltd, and (3) with kind permission of Springer Science and Business Media.

First and foremost I would like to thank my advisor and mentor, Graham Hatfull, for his continued support and guidance throughout my graduate career. From the first day I started in the lab as a rotation student, he has had the dual responsibility of being Chair of the Department and a successful P.I. of a large and demanding lab. But no matter how busy he was, he has always made time for his students. Whenever help was needed, at virtually any time of day, I could always count on his advice for the smallest of science questions or at those times when I felt completely overwhelmed and frantic (which was often). He has a great positive attitude about science and life, and I aspire to be as enthusiastic and knowledgeable as he is. He has made me a better scientist in countless ways, by giving guidance on seminar talks and writing papers, support for scientific conferences, and by providing an encouraging, educational, and fun lab

environment in which to work. His unwavering support of my science career has given me more opportunities than I could have imagined, and I will always be grateful. For all these things (and the continuous beer supply in the breakroom refrigerator), thanks boss!

A number of people have generously given their time, advice, and materials to help me in various ways throughout my project endeavors. I would like to thank Dr. Papavinasasundaram for providing plasmid pKP134, Dr. Richard Kolodner for plasmid pRDK557, and Dr. William Jacobs, Jr. and colleagues for plasmids p0004, p0004S, p0004s:leuB, and p0004S:leuD. I would also like to thank Dr. Jacobs for kindly allowing us to use the avirulent strain of *M. tuberculosis* (*mc*^{2}*7000*), which greatly simplified my recent work with *M. tuberculosis*. I would also like to thank Drs. Lisa Sproul and Troy Krzysiak for their assistance with analytical gel filtration assays. I am also appreciative of Dr. Tony Schwacha and Matthew Bochman for their advice on several biochemical assays, and specifically Matthew Bochman for assistance with electron microscopy and filter binding assays. In addition, I would like to acknowledge Drs. Jeffrey Lawrence and Heather Hendrickson for identification of the mycobacterial *dif* sites and related discussions. Drs. Don Ennis and Gareth Cromie identified the putative mycobacterial AddA proteins, and I would like to thank Dr. Cromie specifically for providing multiple sequence alignments. Finally, Dr. Joanne Flynn and members of her laboratory – in particular, Amy Myers – graciously allowed me to spend a lot of time in their BSL3 lab and gave me excellent technical advice throughout that project, which was all truly appreciated.

Many people have been incredibly supportive by way of discussion and advice. First, I would like to thank Dr. William Jacobs, Jr. for his support and enthusiasm regarding the recombineering project, and for the many discussions we have had over the years about the wonderful things that phages have provided (and in particular, TM4). I would also like to thank

Dr. Donald Court and his colleagues for answering questions related to their *E. coli* recombineering technology, and also for his support at the Molecular Genetics of Bacteria and Phages meeting in 2007. Also, Dr. Kenan Murphy has kindly given his time to answer many of my questions pertaining to the various phage recombination systems. Dr. Keith Derbyshire has also been generous in his gifts of the Δ*recA* (*rec42*) and Δ*recB* strains, as well as his thoughtful insights throughout the development of the mycobacterial recombineering system.

My committee members have been incredibly supportive of me throughout grad school. I have always enjoyed talking to Valerie about everything, from our shared love of sweets and sewing to very helpful discussions about career goals. I hope someday to become as great a teacher as she is, both in the lab and in the classroom. Jeffrey has also been a very helpful and patient teacher, and I particularly appreciate the time he gave to me during comprehensive exams and later on to help me with science questions. Roger is an inspiration, both for his love of science and music, and I thoroughly enjoyed the annual BSO concerts as much as any other science discussion we have had. Bill Jacobs was unbelievably generous in agreeing to share his time and knowledge in the difficult task of being my outside committee member from so far away. I also am grateful to my teaching mentors for all their help and support. Melanie Popa was a joy to work with, always pleasant and enthusiastic, and I learned so much from her. Alison Slinskey-Legg, and all the members of the Gene Team 2006, made me truly appreciate how fun teaching can be. The Gene Team is a wonderful program, and I feel privileged to have been a part of it. I cannot express my thanks enough for all my committee's and mentors' guidance and support over the past five years.

I wish I could take the time to individually thank every person who has made my grad school experience richer, but I fear that would take many pages. I would like to thank the

members of the Hatfull lab, past and present, for all their help over the years. They have given me countless science ideas and materials, listened patiently to my problems, and answered too many questions to count. Each person has helped, no matter how big or small, to make me a better scientist, and I have had a wonderful experience in this lab. I will miss the lab potlucks, cookie exchanges, birthday celebrations, lab dinners at ASM, and even the messy breakroom.

Of all the experiences I have had in my 26 years, I have never had so much fun or met so many amazing people. I want to spend a bit of time to mention some of these people, for they have not only been great friends, but also have become great scientists. I will cherish so many of the close friendships I have made with Becky Gonda, Shruthi Vembar, Maggie Braun, Alycia Bittner, Grace Colletti, Heather Hendrickson, and Stephen Hancock. Many others I will remember for the fun we have had at parties, BASHs, camping, and much more. In particular, I must thank Lori Bibb, who taught me so many of the first things I learned in the Hatfull lab, but also for her enthusiasm and crazy sense of humor that has made me so appreciative of our friendship. Also, I could not have survived grad school without the help, support, and friendship of Laura Marinelli. She has been a wonderful friend to me, both inside and outside the lab, and I am so grateful for everything she has done for me.

Lastly, I would like to thank my family. My mom and dad have always been a constant and steadfast source of support and love, and I feel so lucky to have them. I admire them both for their achievements in their careers and family life, for no one could have had better parents than me. My sisters, Christine and Katie, are both growing up into beautiful, intelligent women. I am proud of them and thankful for their support. Many other family members are in my thoughts, as well, for their love and support, and I thank them. I am also so glad to have become a small part of Matt's family, and I thank all of them for bringing me into their lives. And finally, the newest

member of my family, my best friend, Matthew “Marie” Bochman, deserves so many thanks. He has been so generous with his help in science, but also in so many other countless ways, that this experience would not have been complete without him. I love you all.

1.0 INTRODUCTION

Tuberculosis kills more than one million people each year, and it is estimated that one-third of the global population is currently infected with the causative agent of this disease, *Mycobacterium tuberculosis* [1]. The world struggles to control this epidemic, yet close to ten million new cases are reported each year. Antibiotic resistance in pathogenic bacteria is a continual concern, but it is even more devastating in *M. tuberculosis* when coupled with its persistence. The recent emergence of multiple drug-resistant (MDR) and extensively drug-resistant (XDR) strains of *M. tuberculosis* emphasizes the need for new treatments. Advances in understanding the mechanisms of drug resistance and persistence are therefore critical to improving drug treatments [194].

Scientific study of *M. tuberculosis* requires intricate genetic, molecular, and biochemical approaches to determine what makes it such a successful pathogen. In particular, inactivation of genes by allelic replacement is a crucial first step to understanding gene product function. However, there are several road-blocks to basic genetics in this organism: slow growth, inefficient DNA uptake, and relatively high rates of illegitimate recombination. Although numerous genetic tools have been developed to overcome these limitations, none offer a stream-lined method to manipulate the bacterial chromosome for multiple types of mutagenesis. Clearly there is a need for efficient genetic techniques in *M. tuberculosis*, as well as the other mycobacteria that are studied as model systems.

1.1 GENETICS AND RECOMBINATION IN MYCOBACTERIA

Traditional genetic techniques have been developed and extensively utilized in several genetically tractable bacterial species, such as *Escherichia coli* and *Bacillus subtilis*. For example, to study gene function, gene replacement mutants are often constructed in bacteria using allelic exchange substrates (AESs) that contain homology to a target gene flanking a selectable marker. Homologous recombination facilitates replacement of the endogenous gene with the selectable marker, resulting in a mutant strain. A variety of strategies have been developed for these types of genetic manipulations in several model bacterial systems such as *E. coli*; however, this is not a simple task in most mycobacteria. While some species, such as the non-pathogenic fast-growing species *Mycobacterium smegmatis*, are easier to use for traditional genetics, others like *M. tuberculosis* present huge difficulties for even simple mutagenesis such as allelic gene replacement. This section will examine the obstacles inherent to mycobacterial genetics and the strategies developed to overcome these.

1.1.1 Barriers to genetics in *M. tuberculosis*

M. tuberculosis genetic studies are hindered by two factors related to its growth and cell biology. First, the extremely slow growth rate (>24 hour doubling time) of this bacterium reduces the speed with which any experiments can be performed. Second, the pathogenicity of the organism requires working in a biosafety level three laboratory, which can also be cumbersome and time-consuming. Researchers in this field often turn to other mycobacterial species that are easier to manipulate, such as the fast-growing, non-pathogenic strain *M. smegmatis*, or the avirulent vaccine strain *Mycobacterium bovis* BCG.

While both slow growth and pathogenicity limit the ease and speed with which researchers can manipulate *M. tuberculosis*, there are more specific issues that complicate genetic assays. Because of their waxy coats, mycobacterial cells have a tendency to grow in aggregates or 'clumps' making isolation of single cells for genetic analyses difficult [41]. Additionally, generalized transducing bacteriophages that infect *M. tuberculosis* have not been isolated, and therefore mutations cannot be simply moved to different strain backgrounds as can be done in *M. smegmatis* [111,183].

In the past, the small cache of available antibiotic resistance markers was also a limitation, but this is slowly being overcome [3,19]. Many mycobacterial species encode β-lactamases, which therefore eliminates ampicillin-resistance genes as usable markers, and the instability of tetracycline over time in culture makes it impractical for use with the slow-growing mycobacteria [19]. Other antibiotics, such as chloramphenicol, have been used, but high background resistance make them less desirable [210]. The first demonstration of a selectable marker for the mycobacteria was a kanamycin resistance gene (*aph*; kan^R) used in *E. coli*-mycobacterial shuttle phasmids and on replicating plasmids that enabled stable introduction of foreign genes [62,87,107,210]. Later, hygromycin [59], apramycin [153], streptomycin [78], and gentamicin [121,155] were also successfully utilized. Currently, the kanamycin-resistance (kan^R) and hygromycin-resistance (hyg^R) genes are still the selective markers of choice, although high levels of spontaneous Kan^R colonies are reportedly a problem when using kan^R in some assays [139]. Another method for selection uses the mycobacteriophage L5 gene *71* that confers superinfection immunity such that no antibiotic markers are required, and this is a huge benefit for construction of recombinant vaccine strains [49]. Other selectable markers developed for the mycobacteria include auxotrophic complementation [21] and mercury resistance [16].

Specifically, complementation of strains deleted for auxotrophic genes can be used as a form of selection, which was recently demonstrated with a Δ*leuD M. bovis* BCG strain [21]. Great potential for other selective markers exists from sources, such as mycobacteriophages [70] and mutant alleles isolated from drug-resistant strains.

A low rate of DNA uptake in the mycobacteria has also been troublesome; even the use of electroporation [210] – an improved strategy over spheroplasting [86] – still yields relatively low numbers of transformants of replicating or integrating plasmids in mycobacterial cells. Although protocols for DNA transformation have been optimized repeatedly, typical transformation rates average 10^5 – 10^6 transformed cells per microgram of DNA out of 10^9 viable cells [155,235], even though some have claimed up to 10^7 [139]. The most effective strategy for improving transformation efficiency in *M. tuberculosis* is utilizing warmer temperatures (up to 37°C) during incubations of cells prior to preparation for electroporation. In contrast, lower temperatures (incubating on ice) are preferential for *M. smegmatis* [235]. Further, in comparison to cells that are stored at -80°C prior to use, freshly prepared cells tend to have higher transformation efficiencies [80]. Adding sub-lethal amounts of chemical agents that affect cell wall integrity – such as glycine or ethionamide – can also moderately improve the efficiency of transformation [3,235]. Others have treated the DNA substrates used for allelic replacements with ultra-violet light (UV), alkali, or boiling to increase transformant recovery [80]. Overall, while improvements can be made, transformation of mycobacterial cells will likely never reach the high efficiencies of 10 % (transformants/viable cells) routinely seen in other bacteria such as *E. coli*.

Despite the difficulties described above, the primary obstacle to simple genetics in *M. tuberculosis* is the relatively high level of illegitimate recombination compared to homologous

recombination observed in these bacteria [3,91,125]. During attempts to make targeted gene knockouts in *M. tuberculosis* and *M. bovis* BCG, it was seen that, instead of undergoing homologous recombination with the target locus, linear AESs were incorporated into the genome at seemingly random loci. This occurs at such high frequencies that it prevents simple isolation of a colony that has undergone targeted gene replacement [3,91]. Clearly, illegitimate recombination is a huge impediment to simple genetics in *M. tuberculosis*, and a variety of techniques have been developed to overcome this (see section 1.1.5); the available information on the molecular basis of illegitimate recombination will be examined in section 1.1.3.

1.1.2 Genetics in other mycobacteria

While *M. tuberculosis* is a central focus of research because of the health impact of the disease, other mycobacteria are also commonly studied. There are over 130 species of mycobacteria that have been classified, and these can be characterized broadly as either 'fast-growing' or 'slow-growing,' the latter of which includes the pathogenic species. Many of these are grouped in two classes: the *M. tuberculosis* complex and the *Mycobacterium avium* complex [226]. These are the causative agents of tuberculosis and other diseases in animals and humans, especially in AIDS patients. In addition, although most of the fast-growers are not generally pathogenic, some can cause disease in immunocompromised individuals. Therefore, many mycobacteria are studied as either model systems or as pathogens in their own right. There are inherent characteristics of mycobacteria that make genetic manipulations of these organisms difficult, such as the propensity for cell-clumping and inefficient DNA uptake discussed above. The additional difficulties geneticists encounter with *M. tuberculosis* are also present in other slow-growers: biosafety level three requirements and illegitimate recombination. While there are

innumerable specific differences in manipulations of mycobacterial species, some of the more common model mycobacteria are briefly described below.

The vaccine strain *M. bovis* Bacille Calmette-Guerin (BCG; a member of the *M. tuberculosis* complex) is often used to model *M. tuberculosis* because, even though it is a slow-grower, it is relatively non-pathogenic and can be used in biosafety level two containment. *M. bovis* was passaged 230 times, and the resulting strain has lost the ability to cause disease in several animal models [26]. Experimental evidence has shown that deletion of the Region of Difference 1 (RD1) largely contributes to its attenuation (reviewed in [24]). *M. bovis* BCG exhibits many of the molecular characteristics of *M. tuberculosis*, including limited allelic exchange due to illegitimate recombination [91].

Members of the *M. avium* complex are also frequently studied, including *Mycobacterium intracellulare* and numerous subspecies of *M. avium* [121,226]. Unfortunately, DNA transformation frequencies are particularly low in these organisms, compounded by relatively high levels of inherent antibiotic resistance. However, gene replacement mutants are readily obtained in *M. intracellulare* by homologous recombination, which is unique among the slow-growers [121].

One of the most intractable mycobacterial species is *Mycobacterium leprae*, the causative agent of leprosy, which has never been grown in artificial media and thus is not amenable to classic genetics. However, growth in animal models such as the armadillo and in mouse footpads facilitates metabolic and clinical study of this pathogen [202]. Also, recent sequencing of the *M. leprae* genome has yielded new insights into its genomics and proteomics, enabling better comparisons with more tractable mycobacterial species.

Arguably, *M. smegmatis* is one of the best model mycobacterial species: it grows relatively fast (doubling time of approximately two hours), is non-pathogenic, and is amenable to genetic manipulations [80,82]. Generalized transducing phages that infect *M. smegmatis* have been isolated [111,183], as well as numerous plasmids – both replicating and integrating – and promoter systems that can be used for cloning and gene expression [95,110,126,160,170]. A significant advance to *M. smegmatis* genetics was the isolation of a transformation-proficient strain, mc^2155 [211], with which DNA transformation rates of up to 10^7 colonies per microgram DNA are obtained [139]. The *M. smegmatis* mc^2155 genome has also been sequenced, making this widely-used strain a particularly ideal system.

1.1.3 Recombination in mycobacteria

Attempts at allelic gene replacement in *M. tuberculosis* were unsuccessful initially due to the prevalence of 'illegitimate recombination': recombination between unrelated DNA sequences with very short or no regions of homology [84,125,139]. This type of recombination is observed broadly in prokaryotes and eukaryotes and causes genome rearrangements by two main mechanisms, which are either dependent on, or occur independently, of short homology [50,84,106]. Illegitimate recombination is thought be involved in repair of chromosomal breaks as a mechanism of recombinational repair [106], is induced in response to DNA damage, and is spontaneously induced at lower frequencies [205]. Although it is not surprising that illegitimate recombination occurs in bacteria, the high levels of this found in some mycobacterial species compared to homologous recombination is striking. Illegitimate recombination is troublesome for mycobacterial researchers because it is a barrier to straight-forward genetics.

The relative frequencies of illegitimate and homologous recombination (as assayed by allelic exchange) vary among the mycobacterial species. Although low levels of illegitimate recombination have been reported in *M. smegmatis* [80], sufficient levels of homologous recombination occur, such that gene knockouts are easily obtained [82,125]. In *M. tuberculosis* and *M. bovis* BCG, illegitimate recombination rates are unusually high compared to homologous recombination, making gene replacements difficult to isolate. However, this is not common to all slow-growing mycobacteria, and single homologous recombinants are readily obtained in *M. intracellulare* and *Mycobacterium marinum*. No double crossover events were observed in these studies [80,121,184], indicating that while illegitimate recombination is not frequent in these bacteria, homologous recombination occurs less frequently than in *M. smegmatis* [82]. Overall, genetic manipulation is difficult in most slow-growing pathogenic mycobacteria; some of the initial experiments illustrating this are discussed below.

1.1.3.1 Gene replacement by homologous recombination in *M. smegmatis*

The first report of successful targeted gene replacement in *M. smegmatis* was accomplished by using a 'suicide vector' [82], which is a plasmid that replicates in *E. coli* for propagation but lacks a mycobacterial origin of replication and relies on a homologous recombination event to integrate into the mycobacterial chromosome (see section 1.1.5). The plasmid was constructed with a kan^R gene flanked by DNA segments with homology to the *M. smegmatis pyrF* gene. This locus was chosen because strains with a wild type *pyrF* gene can grow in media without uracil but are inviable in the presence of 5-fluroorotic acid (5-FOA), while Δ*pyrF* strains are uracil auxotrophs and 5-FOA resistant. These characteristics, therefore, provide both positive and negative selection, and single versus double homologous recombination events can be distinguished (Figure 1). Using this approach, single and double

crossovers occurred at similar frequencies in *M. smegmatis* (60% and 40%, respectively), although these frequencies vary in other reports [168,196]. Gene replacement also occurs when a second crossover event loops out the remaining vector sequence from the first single crossover. These mutants can be identified by selection with 5-FOA and arise at a frequency of 10^{-4} (Figure 1B). Other groups have developed similar strategies for constructing gene replacements in *M. smegmatis*, some of which include the use of other counter-selectable markers that make double crossover allele identification easier [168,196]. Compiling data from multiple studies, frequencies of homologous recombination resulting in plasmid integration (single crossovers) average 10^{-3} – 10^{-4} cfu per microgram DNA, with respect to the number of colonies that arise from transformation with a replicating control plasmid [168,196]. Gene replacement events (double crossovers) are less frequent but still occur at a frequency of 10^{-4} – 10^{-6}, which makes *M. smegmatis* an ideal model system for mycobacterial genetics [125].

Figure 1. Homologous recombination in *M. smegmatis*.

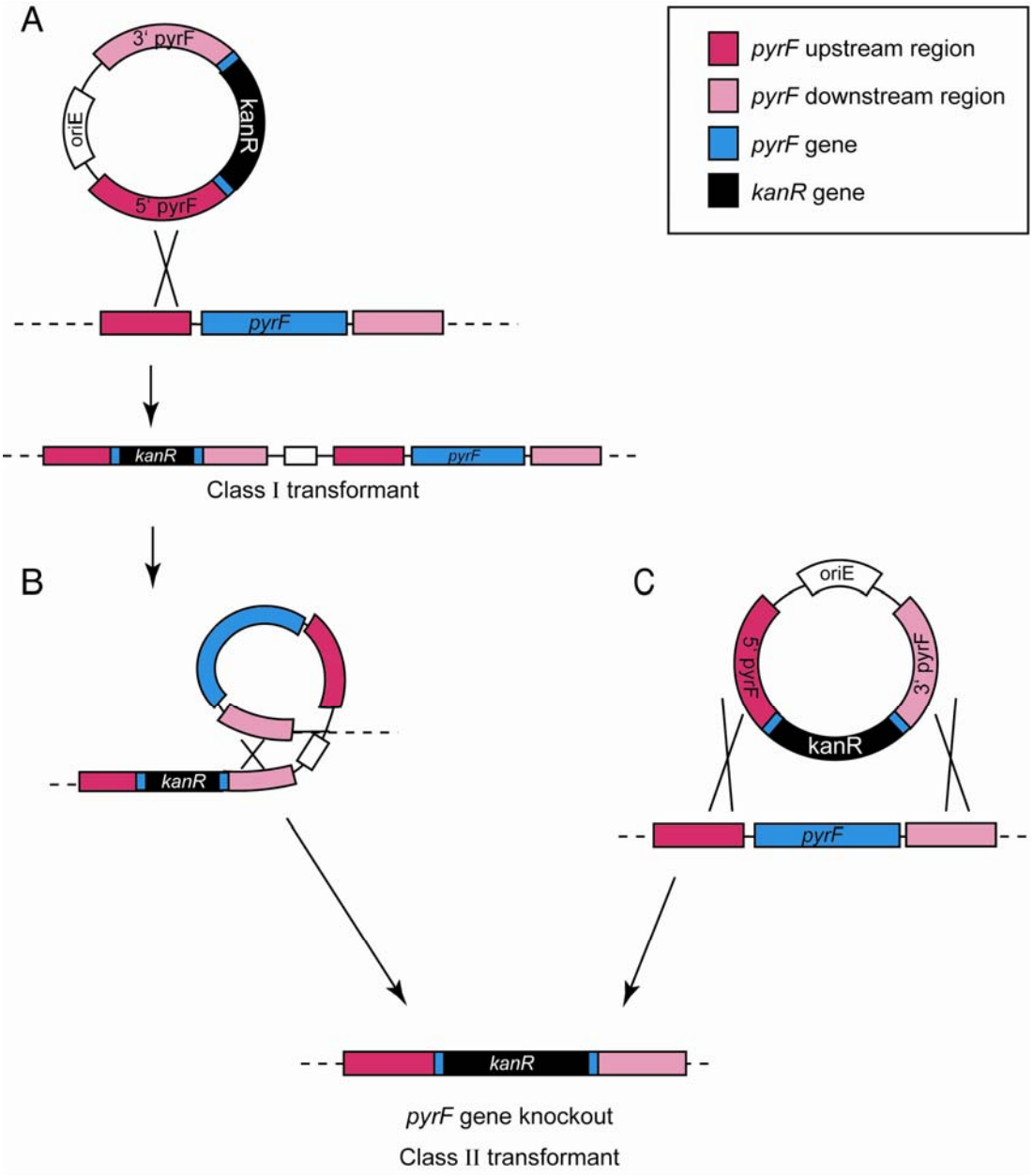


Figure 1. Schematic representing the classic allelic gene replacement experiments targeting the *pyrF* gene performed by Husson *et al.* [82] using a circular suicide vector. (A) Class I transformants: a single homologous recombination event yields an integration of the entire plasmid; transformants are Kan^R, uracil prototrophs, 5-FOA sensitive. (B) Following a single crossover, the sequences that are in duplicate can be removed by a second recombination event. (C) Class II transformants: a double homologous recombination event yields integration only of the kan^R gene into the *pyrF* gene; transformants are Kan^R, uracil auxotrophs, 5-FOA resistant.

1.1.3.2 Evidence of illegitimate recombination in *M. tuberculosis*

In the first report of illegitimate recombination in the mycobacteria, Kalpana *et al.* were unable to replace either the *M. bovis* BCG or *M. tuberculosis* strain H37Rv *met* genes with a *kan*^{R} gene [91]. No correctly targeted gene replacements were identified out of more than 200 Kan^R colonies screened (Table 1). Linear double-stranded DNA (dsDNA) AESs were used in an attempt to exclusively isolate mutants from double crossovers, and this resulted in an ~10-fold increase in colonies. However, Kan^R recombinants were recovered irrespective of the presence of homologous mycobacterial DNA sequences in the AES (pBR322::Tn5, Table 1), clearly showing that integration of the *kan*^{R} gene was not dependent on homologous DNA sequences. The illegitimate recombinants were obtained at a relatively high frequency (*i.e.* 10^{-4} to 10^{-5} relative to plasmid transformants), such that they masked the presence of colonies (if any) arising from correctly targeted recombination events. In the same study, *M. smegmatis met* mutants were readily obtained using either linear or circular AESs, as expected from previous studies [82], and recovery of recombinants was dependent on the presence of mycobacterial sequences in the AES (Table 1). Subsequently, other groups successfully isolated mutant alleles that were generated by homologous recombination in *M. tuberculosis*, but the frequencies of single crossovers were low (<20%) and were lower for double crossovers (<5%) [3,8,147,168,188].

Table 1. Isolation of illegitimate recombinants in *M. tuberculosis* and *M. bovis* BCG.

Plasmid	DNA type[d]	*M. smegmatis*	*M. bovis* BCG	*M. tuberculosis*
No DNA control[a]	–	0	12	7
pYUB53 control (per μg DNA)[b]	CCC (replicating)	10^6	1-3 x 10^5	3 x 10^4
pBR322::Tn5[c]	CCC	0	13	–
	Linear	0	140	26
M. smeg met::Tn5	CCC	332	–	–
	Linear	196	–	–
M. bovis met::Tn5-a	CCC	–	16	–
	Linear	–	130	–
M. bovis met::Tn5-b	CCC	–	18	14
	Linear	–	148	27

Summary of data from DNA transformations performed by Kalpana *et al.* targeting the *met* genes of *M. smegmatis*, *M. bovis* BCG, and *M. tuberculosis* [91].
a. The 'no DNA control' determined background Kan^R.
b. pYUB53 is an episomally replicating plasmid used to determine overall transformation proficiency (per μg).
c. Plasmid pBR322::Tn5 contains no mycobacterial sequences and is a control for illegitimate recombination.
d. Cells were transformed with 2-4 μg AES DNA, either covalently closed circular (CCC) or linearized, containing a Tn5 *seq1* inactivated *met* gene; transformants were selected on Kan and the number of colonies are reported.

1.1.3.3 The recombination genes of *M. tuberculosis*

The complete genome sequence of *M. tuberculosis* has made comparative genomic studies possible [35], which revealed genes predicted to encode homologous recombination proteins [130]. The *M. tuberculosis* predicted open reading frames (ORFs) were searched for homologues of the *E. coli* recombination (rec) proteins, and a number of predicted rec proteins were identified including RecA, RecBCD, RecF, RecR, as well as Holliday junction resolvases [130]. Strikingly, several rec proteins were not identified in this initial analysis, such as RecO, RecJ, ExoI, RecQ, SbcCD, and RecET.

However, closer analysis indicates that identification of mycobacterial recombination proteins cannot be identified merely through the presence of *E. coli* rec homologues. In fact, many mycobacterial species, including *M. tuberculosis*, do have ORFs that encode proteins with similarity to these 'missing' rec proteins [35]. The *M. tuberculosis* RecO protein is easily identifiable by BLAST analysis but is only distantly related to the *E. coli* RecO. In addition, *M.*

tuberculosis Rv2837c is a member of the DHH protein family, which includes RecJ proteins from several bacteria including *E. coli.* Another *M. tuberculosis* ORF, Rv3198c, is predicted to encode a protein that has both a UvrD2 helicase domain and a fragment of the RecQ domain, and is therefore described as a putative RecQ helicase. Finally, since the RecET proteins are encoded by a cryptic prophage in *E. coli*, it is not surprising that these are absent in *M. tuberculosis*. Therefore, this bacterium has a number of recognizable recombinational repair pathway components.

Arguably, a comparison of the known recombination genes of more closely related bacteria may provide better insights into the recombinational repair system of *M. tuberculosis*. Comparative analysis of the *B. subtilis* and mycobacterial genomes revealed the presence of multiple genes encoding *B. subtilis* AddA homologues (at least two) in several mycobacterial species, including *M. smegmatis* and *M. tuberculosis* (D. Ennis and G. Cromie, personal communication; see Appendix A and Figure 34). The AddAB proteins function similarly to RecBCD for processing and repair of dsDNA lesions and are most commonly found in Gram-positive bacteria, whereas RecBCD are typically encoded by Gram-negative bacteria [32,245]. The specific activities of RecBCD have not been fully characterized in mycobacteria for general recombinational repair, and these have only been examined with regard to their role (or lack thereof) in conjugation and non-homologous end-joining, respectively [120,234]. It is therefore possible that both sets of recombination proteins – RecBCD and the two AddA homologues – are active and perhaps redundant in mycobacteria. Alternatively, it may be that only one set of proteins is expressed and/or active *in vivo*.

1.1.3.4 The debate over homologous and illegitimate recombination in mycobacteria

It was not clear from the initial studies discussed above if levels of homologous recombination are actually decreased or if the levels of illegitimate recombination are merely increased – or perhaps both – in slow-growing mycobacteria such as *M. tuberculosis*. One hypothesis is that the presence of an intein in the *M. tuberculosis recA* gene reduces the activity of this pivotal recombination enzyme, thereby decreasing overall levels of homologous recombination (reviewed in McFadden, 1996).

In *M. tuberculosis*, the conserved RecA sequences are situated at the N- and C-termini of the ORF and are interrupted by 440 amino acids that are not conserved in other RecA proteins [45]. Splicing of the full-length protein is essential to remove this "spacer protein," and the N- and C-terminal regions are ligated to produce the mature active protein [46]. The *recA* gene of *M. leprae* also includes an intein that is spliced *in vivo* [57], but the *recA* gene of *M. smegmatis* does not [47], which further suggests that the abnormal gene structure of the *M. tuberculosis recA* may correlate to low levels of homologous recombination. *In vitro* experiments with purified *M. tuberculosis* RecA proteins – both full-length and mature – have shown that the unspliced protein is defective in ATPase activity and strand exchange, whereas the mature protein is active [103]. It is therefore possible that RecA activity *in vivo* is regulated by conditional splicing of the full-length inactive protein.

In addition, expression of *recA* in *M. tuberculosis* is controlled by multiple transcriptional regulatory elements, which adds to the complexity of regulation. Two promoters upstream of *recA* are regulated in response to DNA damage, one by LexA and RecA in the classical mechanism through an SOS box, while the other is independent of LexA and RecA (discussed below) [48,65,127]. Additionally, RecA activity is negatively regulated by a co-transcribed

protein RecX in mycobacteria [154,155,230]. It is also intriguing that *recA* expression is much more delayed in response to DNA damaging agents in *M. tuberculosis* as compared to *M. smegmatis* [127,156]. It was suggested, therefore, that the genetic and biochemical characteristics of *M. tuberculosis* RecA may result in reduced levels of homologous recombination in this bacterium.

Subsequent experiments, however, suggested that the intein does not affect the function of RecA in recombination or other activities. Expression of the *M. tuberculosis* RecA – with or without the intein – in an *M. smegmatis* Δ*recA* strain was sufficient to promote levels of homologous recombination similar to wild type *M. smegmatis*, and no illegitimate recombination was observed [56,155]. These data support two conclusions: 1) the *M. tuberculosis* RecA protein inteins does not reduce the levels of homologous recombination in *M. smegmatis*, and 2) the expression of *M. tuberculosis* RecA in *M. smegmatis* is not sufficient to introduce levels of illegitimate recombination similar to those in *M. tuberculosis*. However, similar experiments expressing the *M. smegmatis recA* in an *M. tuberculosis* Δ*recA* strain would be required to determine the specific role of RecA in illegitimate recombination. It is also possible that there are factors regulating RecA splicing in *M. tuberculosis* that modulate its recombination activity levels, and perhaps this does not occur in *M. smegmatis*.

There is evidence that suggests that the levels of homologous recombination are not decreased in *M. tuberculosis*. Experiments by Pavelka *et al.* showed that similar numbers of homologous transformants were obtained in *M. smegmatis*, *M. tuberculosis*, and *M. bovis* BCG using circular suicide vectors, suggesting that illegitimate recombination likely occurs predominantly with linear DNA substrates [163]. These data imply that homologous

recombination frequencies in mycobacteria are similar, and the increased level of illegitimate recombination is likely what is different between the fast- and slow-growing mycobacteria.

It has also been speculated that the slow induction of *recA* expression in *M. tuberculosis* may result in deficiencies in DNA repair and decreased SOS response, leading to high rates of illegitimate recombination. Since *recA* expression is induced slowly (compared to *M. smegmatis*) in response to DNA damage, this could result in reduced RecA-dependent autocatalytic cleavage of LexA and decreased activation of downstream genes involved in the SOS response. In this situation, it is conceivable that chromosomal breaks would be more prevalent, perhaps leading to higher rates of illegitimate recombination for repair of these lesions. The LexA protein of *M. tuberculosis* has been characterized and shown to bind an SOS box (as is typically seen with this repressor [127,128]), and one SOS box is present in one of the promoter regions at *recA*. However, it was found that two mechanisms for DNA damage response exist in *M. tuberculosis*, one that is classically dependent on RecA and LexA and one that is independent of this process; each mechanism controls a different set of genes [48,127,186]. Therefore it seems that even though induction of *recA* expression is slow, other mechanisms for DNA repair and SOS response are in place, perhaps negating the argument that *recA* expression kinetics play a role in illegitimate recombination. Thus, the molecular basis of the relatively high frequencies of illegitimate recombination in *M. tuberculosis* and other slow-growing mycobacteria remains an open question.

1.1.4 Mycobacteriophage-derived genetic tools

Bacteriophages have long demonstrated their utility as sources for genetic tools in bacterial model systems, especially those that are genetically intractable. Over fifty mycobacteriophages

have been isolated and sequenced to date ([73,165] and unpublished data), from which a plethora of genetic information has been gathered, enabling the study of numerous phage genes [71,72]. For the mycobacteria, phage-derived vectors have proven extremely useful for expression of foreign genes. Several integration-proficient vectors containing phage integration cassettes have been developed and can be used simultaneously for stable introduction of multiple genetic elements in a single cell [95,111,126,170]. Also of great use are shuttle phasmids, which are chimeric cosmid molecules containing mycobacteriophage and *E. coli* plasmid DNA [86]. These replicate as plasmids in *E. coli* and as phages in mycobacteria and are used as delivery vehicles; their use for delivering AESs will be discussed in further detail in section 1.1.5.5 [14]. Shuttle phasmids have also been used to deliver transposons for genetic assays [13] and as reporter phages in clinical studies to assay for live mycobacterial cells and drug susceptibility [10,27,88,164,189,197].

Phages have also been isolated that infect *M. smegmatis* and facilitate generalized transduction, enabling transfer of mutations to other strains [111,183]. Generalized transduction would be particularly useful for studying mutations conferring drug-resistance. However, no generalized transducing phages that infect the slow-growing mycobacteria, such as *M. tuberculosis*, have been isolated. Also of use in *M. tuberculosis* are phage-derived methods for selection that can be used in place of antibiotic markers, which are not desireable in potential vaccine strains. The mycobacteriophage L5 repressor gene product gp71 confers immunity to superinfection. Thus, when gene *71* is expressed as a selective marker on plasmids, cells are resistant to infection by a homo-immune phage [49]. Phage promoters have also been used for gene expression in mycobacteria as an alternative to constitutive strong promoters such as the *M. bovis* BCG *hsp60* promoter [18,72]. It is clear that mycobacteriophages have contributed greatly

to the study of genetics in mycobacteria and will likely continue to do so as we learn more through isolation and characterization [73].

1.1.5 Genetic techniques for allelic replacement

Characterization of isogenic mutants is a powerful method for the study of gene function, and targeted gene replacement is a standard way to construct these defined mutants. Other techniques such as transposon mutagenesis and random mutagenesis are extremely valuable but do not offer the same precision or control over the type of mutations made. In many organisms, allelic gene replacement is simple and fast, requiring little DNA manipulation and screening [38]; however, this is not the case for the mycobacteria. Canonical substrates for targeted gene replacement (AESs) contain a selectable genetic marker flanked by long (>1000 bp) regions of homology to the gene locus being targeted. These substrates are introduced into the cell and homologous recombination leads to single or double crossovers to yield a marked allelic replacement mutant. While this strategy is successful in *M. smegmatis*, the prevalence of illegitimate recombination in some of the slow-growing mycobacteria prevents this from being an efficient method for gene replacement. Null mutations in genes resulting in an auxotrophic or otherwise identifiable phenotype were the first constructed because they facilitated differentiation of double versus single crossovers [3,8,9,80,158,188]. Clearly not all gene mutants would have screenable phenotypes, and therefore even the limited success of these early methods suggested a need for improvement.

A number of attempts have been made to improve the recovery of mutant alleles from double homologous recombination events and reduce the need for screening. Figure 2 summarizes the multitude of techniques that were developed for the mycobacteria in a timeline

style and also shows the first gene replacements made in some of the more commonly studied mycobacteria. The majority of mycobacterial genetic tools developed were aimed at modifying the AES to make it more recombinogenic: altering the structure, treatments prior to transformation, and delivery method. The preferred genetic techniques are successful because they either utilize a selection for double crossovers or drastically reduce or eliminate illegitimate recombination events. It is worth noting, however, that none of the strategies developed thus far have successfully increased the levels of homologous recombination in *M. tuberculosis*. This may be due to the complexity of recombination in the mycobacteria, or perhaps this was attempted and never accomplished. Yet this still represents another potential method for improving recovery of allelic replacement mutants.

Figure 2. Development of allelic gene replacement techniques in the mycobacteria: 1990-present.

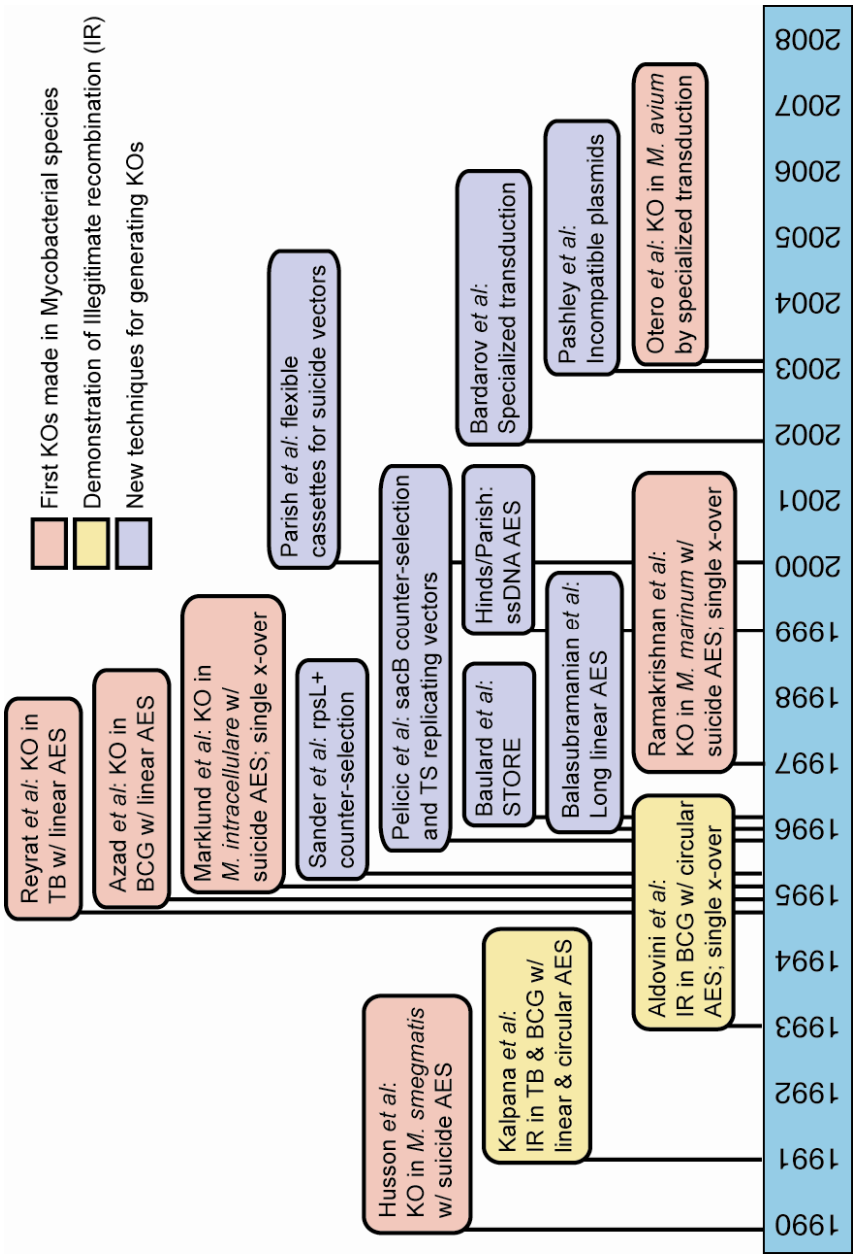


Figure 2. The first gene replacements made in *M. smegmatis*, *M. bovis* BCG (BCG), *M. tuberculosis* (TB), *M. intracellulare*, *M. marinum*, and *M. avium* are indicated by red boxes. The first publications that studied illegitimate recombination (IR) through gene replacement are shown in orange. New techniques are shown in purple boxes. Abbreviations: TB: *M. tuberculosis*; KO: gene knockout, x-over: crossover; STORE: selection technique of recombination events.

Arguably, there were two techniques that were most successful: (1) the use of suicide vectors with counter-selectable markers, which aid in the selection of the desired double-crossover events, and (2) the delivery of the AES by mycobacteriophages (referred to as 'specialized transduction'). This section will discuss the numerous genetic tools developed for the mycobacteria over the last 18 years.

1.1.5.1 AES structural modifications

Numerous AES designs were explored to optimize allelic exchange frequencies: linear versus circular DNA substrates, the length of sequence identity, the presence of nonhomologous DNA flanking the homologous regions, and the selectable marker. The initial experiments performed by Kalpana *et al.* used both a linear and circular dsDNA AES [91], while Aldovini *et al.* used a circular suicide vector as an AES [3]. Using a linearized AES yielded up to ten-fold more colonies than the circular substrate and resulted in mostly illegitimate events in multiple studies [91,163]. It therefore appears from these experiments that: (1) using a circular AES yields lower numbers of recombinants compared to a linear AES, but these result from predominantly illegitimate recombination and single crossover events in *M. bovis* BCG and *M. tuberculosis* [3,91], (2) using linear AESs did not result in any identified homologous recombination events (single or double crossovers), only illegitimate events [91] in *M. tuberculosis* and *M. bovis* BCG, and (3) using circular AESs in *M. smegmatis* can facilitate both single and double homologous recombination events [82] with low amounts of illegitimate recombination [80]. Later experiments with linearized AESs were somewhat successful in *M. tuberculosis* and *M. bovis* BCG for making double crossover mutants, although at low frequencies (~4%) [8,188].

Balasubramanian *et al.* succeeded in making gene replacements in leucine biosynthetic genes using long (40-50 kbp) linear AESs [9]. Genomic cosmid libraries of *M. tuberculosis*

H3Rv and *M. bovis* BCG were constructed, and interplasmid recombination in *E. coli* was used to make the *kan*^{R}-marked disrupted *leuD* allele. In this case, transformants were obtained equally with linear or circular cosmid AESs, but leucine auxotrophs were only found with the linear AES; 6% double crossover mutants were identified. While this was a successful method, it was time-consuming, and another group demonstrated similar frequencies (4%) of double crossover using linear AESs with short (>1 kbp) homologies [188], albeit at a different locus.

Since low levels of spontaneous resistance to kanamycin occur in slow-growers [91], others have used different antibiotic resistance genes such as *hyg*^{R}, gentamicin resistance (*gent*^{R}), streptomycin resistance (*str*^{R}) and even mercury resistance as markers [14,15,82,147,159,161]. However, these methods did not generally improve the recovery of double crossover mutants. It was also suggested that the presence of nonhomologous sequences flanking the homology targeting the gene might increase the propensity for the AES to undergo illegitimate recombination [3,91], although this has not been tested rigorously.

1.1.5.2 Treatment of the AES

Neil Stoker's group has shown that treating the DNA substrate with agents that promote the formation of single-stranded DNA (ssDNA) improves the frequency of homologous recombination in *M. smegmatis*, *M. intracellulare*, and *M. tuberculosis* [80,158]. The most effective experiments utilized treatments with alkali or by boiling to denature the DNA, or merely used ssDNA derived from phagemids. In experiments with ssDNA AESs, not only were transformant numbers typically increased, but also the proportion that had undergone double crossovers. Importantly, the use of phagemid DNA eliminated the recovery of illegitimate transformants.

1.1.5.3 Plasmid delivery of the AES

Numerous groups have also made allelic exchange mutants in mycobacteria using either a circular or linearized suicide vector [3,8,121,159,184,188]. These are plasmids that rely on integration via homologous recombination for maintenance in the mycobacteria, either through a single crossover (in which the entire plasmid is integrated) or double crossover (in which the targeted chromosomal gene is replaced by the disrupted gene) (see Figure 1). Despite the high frequency of illegitimate recombination in the slow-growers, homologous recombination using these substrates is still relatively successful. Further, although single crossovers occur at a higher frequency than double crossovers, single crossover mutants can be propagated and screened for a second recombination event between the duplicate sequences to loop out the excess vector (Figure 1); however, this does not occur at a high frequency [91]. Plasmids with multiple cloning sites flanking different antibiotic markers were constructed to simplify synthesis of the AES suicide plasmid [159], but the screening was still labor-intensive. The development of a two-step counter-selection strategy (discussed below) greatly improved this by reducing the number of transformants screened.

Since the frequency of homologous recombination is lower than the transformation rate in mycobacteria, large quantities of DNA are required for transformations (up to 4 μg). The use of a replicating vector for delivery of the AES could arguably work better than a suicide vector, since extended survival of the plasmid would likely improve the frequency of recombination with the target. A replicating plasmid was used in one study, but did not result in a stable mutant allele of the targeted gene *accBC* in *M. bovis* BCG. However, the reason for this is unknown since PCR and Southern blot analysis confirmed that homologous recombination with the AES had occurred [147]. Another group developed a technique called STORE (selection technique of

recombination events) that uses a replicating plasmid with a promoter-less *kan*^{R} gene targeted to the *M. bovis* BCG *hsp60* locus for replacement of the *hsp60* gene [15]. Selection for Kan^R therefore yielded recombinants that had undergone homologous recombination at the *hsp60* locus, which placed the *kan*^{R} gene under control of the constitutive *hsp60* promoter. However, extension of this technology for targeting other loci would require that the gene is expressed.

One concern with replicating vectors is removing the plasmid; temperature-sensitive plasmids offer an advantage here, but for best results in the slow-growing mycobacteria these are combined with SacB counter-selection (examined in more detail below) [169]. Pashley *et al.* made use of incompatible plasmids to facilitate removal of the plasmid following gene replacement [161]. This technique uses a pair of plasmids that replicate co-dependently and are lost in the absence of selection. The plasmid carrying the AES can therefore undergo targeted gene replacement. However, this method like many others requires multiple rounds of selection, growth, and plating, making it less efficient than other techniques.

1.1.5.4 The counter-selection strategy

Husson *et al.* was the first to use counter-selection for allelic exchange in the mycobacteria (discussed in section 1.1.3.1). In this study, the *pyrF* gene in *M. smegmatis* was replaced with a *kan*^{R} gene through a double crossover event. The mutant was selected by plating on 5-FOA, since loss of wild type *pyrF* confers resistance [82]. This technique was extended later by Knipfer *et al.* who used the *pyrF* gene as a selective marker in a Δ*pyrF* strain for unmarked introduction of genes [97]. Since this is therefore limited to the *pyrF* locus, broader strategies were developed. Another useful counter-selection strategy is the introduction of the wild type *rpsL* gene (*rpsL*+) in a strain with a specific *rpsL* mutation that confers streptomycin resistance (Str^R) [196]. Plating a strain that contains both wild type and mutant alleles on

streptomycin selects for loss of the wild type *rpsL* gene. Therefore when *rpsL+* is placed on a suicide AES, double selection on streptomycin and kanamycin (*e.g.*, if kan^R is the disrupting genetic marker) results in generation of predominantly double crossover gene replacement mutants in *M. bovis* BCG. However, this requires the use of a Str^R resistant strain background, which is not ideal for vaccine development.

The *B. subtilits sacB* gene has been extremely useful as a counter-selective marker in mycobacterial genetics. The presence of the *sacB* gene causes sensitivity to sucrose, and therefore plating on sucrose selects for loss or mutation of the gene (Figure 3) [166-168]. Allelic exchange mutants that are the products of double homologous recombination events can be obtained in a single step by dual positive and negative selection with antibiotics and sucrose at 100% efficiency. Alternatively, if this is unsuccessful, allelic exchange can be performed in two steps, in which single crossover mutants are selected by antibiotic resistance, followed by removal of the vector sequence by a second crossover event, selected by plating on sucrose (this occurs in ~two-thirds of the colonies screened). This strategy can also be used to make unmarked mutants; in this case, γδ resolvase sites are placed flanking the antibiotic marker and *sacB*, and recombinants from expression of the resolvase can be selected by sucrose resistance (Figure 3). The *sacB* gene has also been used for very effective gene replacement (100%) on replicating temperature-sensitive plasmids as AES delivery vehicles: it ensures loss of the plasmid by shifting to high temperature and plating on media with sucrose [169].

Figure 3. Gene replacement by counter-selection with *sacB*.

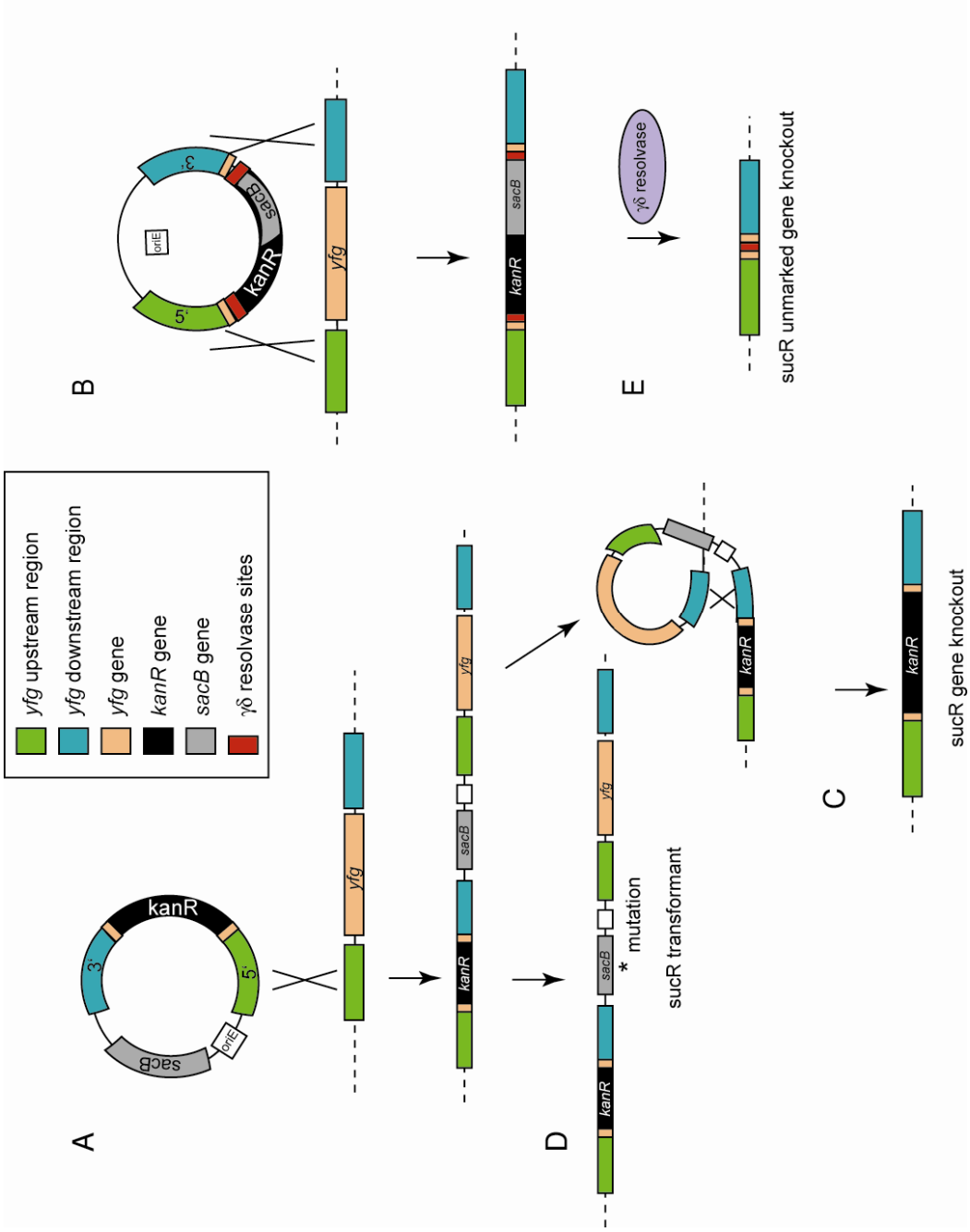


Figure 3. Genes (yfg: your favorite gene) targeted by using *sacB* on the vector DNA result first in a (A) single crossover and then loop out the vector, or a (B) double crossover in vectors which contain the *sacB* gene on the backbone. Simultaneous selection for antibiotic resistance (*e.g.* Kan^R) and sucrose resistance can yield either (C) removal of the vector containing sacB or (D) mutation of sacB. (E) Unmarked mutations can also be generated by using γδ resolvase: res sites are placed flanking the antibiotic resistance gene (e.g. kan^R) and *sacB* (instead of it being on the vector backbone).

1.1.5.5 Specialized transduction

Delivery of the AES by phage infection, a method called 'specialized transduction,' has proven to be a successful method for targeted gene replacement [14]. This was accomplished by the development of shuttle phasmids, which are chimeric DNA molecules that replicate as plasmids in *E. coli* and phages in mycobacteria [86]. Phasmids contain phage genomic DNA with an *E. coli* plasmid inserted in a non-essential region of the genome (Figure 4). They can therefore replicate as plasmids in *E. coli* and as phages in mycobacteria. This technology was developed by Jacobs *et al.* using mycobacteriophage TM4, and later mycobacteriophages D29 and L1 [86,164,210]. The most commonly used shuttle phasmid is phAE87, which is a TM4 shuttle phasmid containing a temperature-sensitive mutation that allows phage propagation at 30°C but not at 37°C [13]. Shuttle phasmids have been used not only for delivery of transposons and expression of reporter genes, but also for delivery of AES for targeted gene replacement in both the fast- and slow-growing mycobacteria [13,14,88,164].

Figure 4. Construction of TM4 shuttle phasmids.

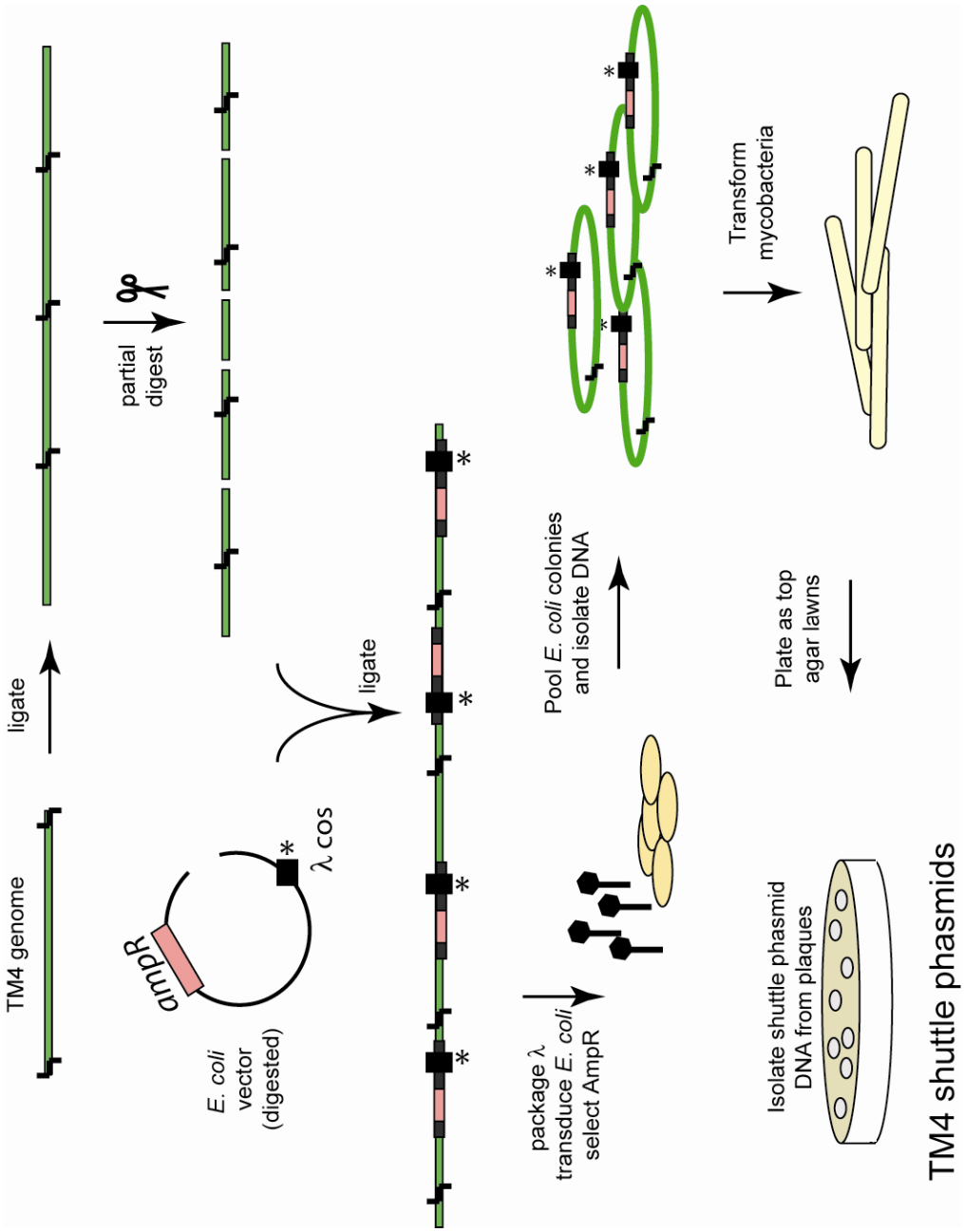


Figure 4. Construction of the parent shuttle phasmid. Phage DNA is ligated together via the sticky ends of the genome to form concatemers, and these are partially digested with a frequently-cutting restriction enzyme (such as *Sau3A*I) to cut minimally in the genome. Fragments ~45 kbp in length are ligated to an *E. coli* vector (digested with an enzyme leaving a compatible site) that contains a phage λ *cos* site for packaging and an ampicillin resistance gene (amp^R). These molecules are packaged into λ phage heads *in vitro*, *E. coli* cells transduced, and colonies are selected on ampicillin. Pools of *E. coli* colonies are made and DNA isolated; this is transformed into mycobacteria and cells are plated as top agar lawns. DNA constructs that form plaques and retain the *E. coli* plasmid are true shuttle phasmids.

For gene replacements, a canonical AES is constructed by cloning ~1000 bp of upstream and downstream homology to the target gene flanking an antibiotic marker (typically *kan*^{R} or *hyg*^{R}). This can be directly cloned into a parent shuttle phasmid such as phAE87 to replace the existing *E. coli* plasmid sequences, and shuttle phasmid molecules containing the AES are prepared. A mycobacterial culture is then infected with mycobacteriophage-packaged shuttle phasmids at a non-permissive temperature for phasmid replication, and this facilitates delivery of the AES and targeted gene replacement (Figure 5). This method has been used to make more than 300 gene mutants in *M. tuberculosis* (W.R. Jacobs, Jr., personal communication). Specialized transduction has also been used to construct a strain of *M. tuberculosis* containing a single defined point mutation in the *inhA* gene [232]. This was the first experiment in which a point mutation was placed in an endogenous gene in a wild type background, and is an example of the power of specialized transduction.

Figure 5. Targeted gene replacement by specialized transduction.

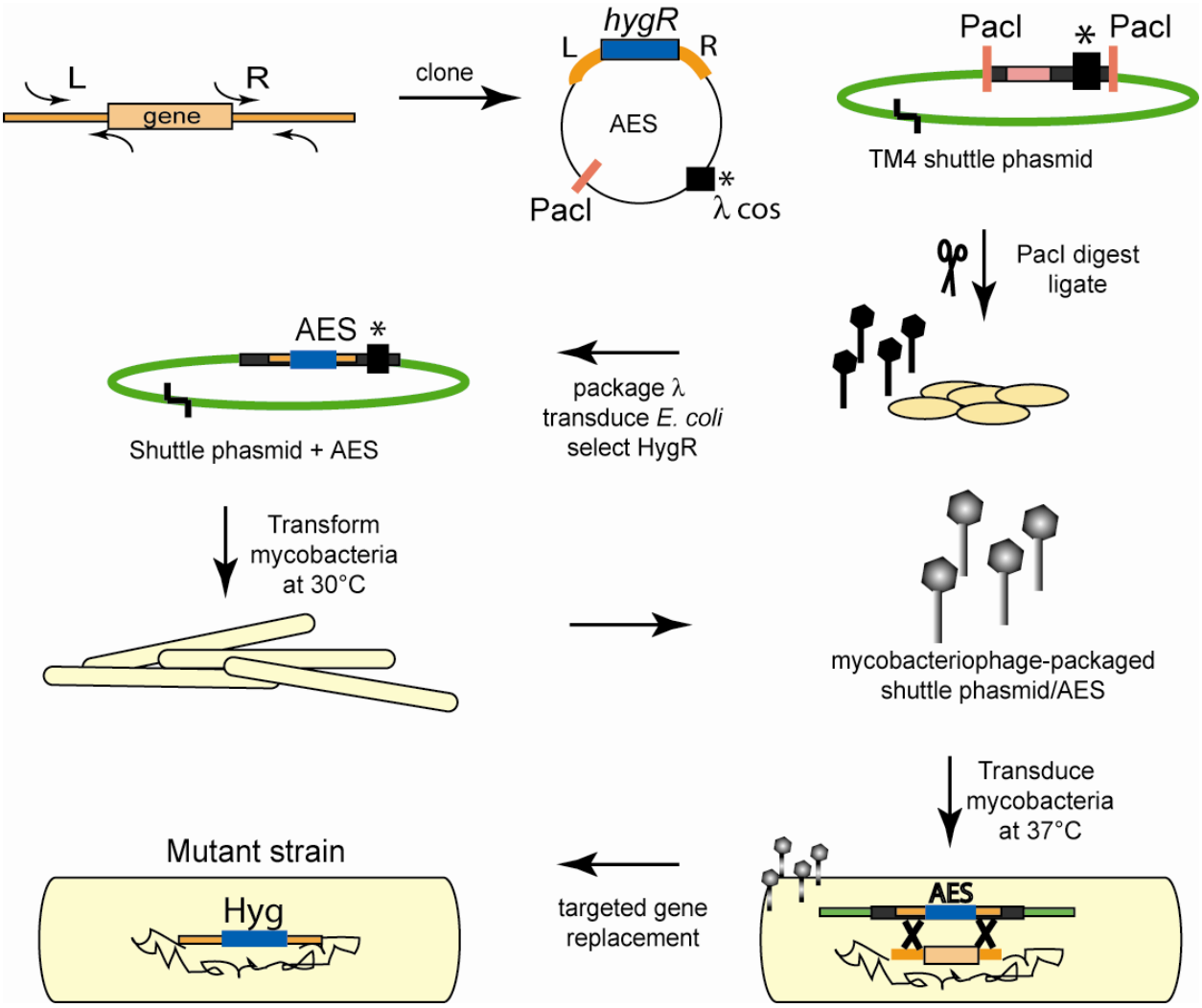


Figure 5. Upstream and downstream regions of the target gene are cloned flanking an antibiotic resistance gene (e.g. *hyg*^{R}). Shuttle phasmids for gene replacements are then constructed by using the parent shuttle phasmid (such as phAE87) and inserting the AES vector by restriction digest with *Pac I* and ligation. These are packaged into λ heads, *E. coli* infected and Hyg^{R} colonies selected. The shuttle phasmid DNA is prepared and transformed into mycobacteria at permissive temperature (30°C) and resulting plaques are picked and lysates of phage prepared. Mycobacteria are then transduced with the phage at a non-permissive temperature (37°C) and the AES will undergo homologous recombination with the target in the genome yielding a gene replacement mutant.

In conclusion, a variety of techniques have been developed for gene replacement mutagenesis of *M. tuberculosis* with varying success. Each method has drawbacks that include time-consuming AES constructions or screening of large numbers of recombinant colonies. In other organisms, technologies for mutagenesis have been greatly improved through the use of phage-encoded recombination proteins. In particular, genetics in *E. coli* and related Gram-negative bacteria have benefited enormously by exploiting these recombination proteins in a genetic system called recombineering. The following sections will discuss the recombination proteins of bacteriophages that promote single strand annealing homologous recombination and their use for development of host genetic tools.

1.2 SINGLE STRAND ANNEALING PROTEINS

Homologous and non-homologous recombinational repair of DNA is an extremely well-studied field that is exemplified by research in *E. coli* and bacteriophage λ [106]. Homologous recombination – the pairing and exchange of complementary strands – can be divided into two mechanisms: strand invasion and single strand annealing. The two classically defined mechanisms of RecA-dependent strand invasion are: the 'daughter strand gap repair pathway' involving the RecF 'machine,' and the 'double-strand end repair pathway' mediated by the RecBC complex (reviewed in Kuzminov 1999). Although alternative repair pathways exist that involve different combinations of the Rec proteins, it is clear that RecA plays a central role in recombinational repair of chromosomal lesions that occur during replication and DNA damage.

The second major recombination pathway that appears to be conserved through eukaryotes is called the 'single strand annealing pathway.' As the name implies, single strand

annealing involves pairing of complementary single strands via a RecA-independent mechanism that is initiated at double strand breaks (Figure 6) [220]. These recombination proteins, called single strand annealing proteins (SSAPs), promote strand pairing, strand exchange, and strand invasion [17,69,114,129,145,193]. SSAPs are found predominantly in bacteriophages and in bacterial genomes in prophages, although they have also been identified in eukaryotes, including yeast and humans. The SSAPs comprise three superfamilies based on sequence conservation: (1) the Red β/RecT family, (2) the Erf family, and (3) the Rad52 family [85]. It appears that these all have bacteriophage origins and are typically found adjacent to other DNA recombination or repair proteins, such as exonucleases. These groups of proteins and their biochemical characteristics will be explored in this section.

Figure 6. Single strand annealing pathways.

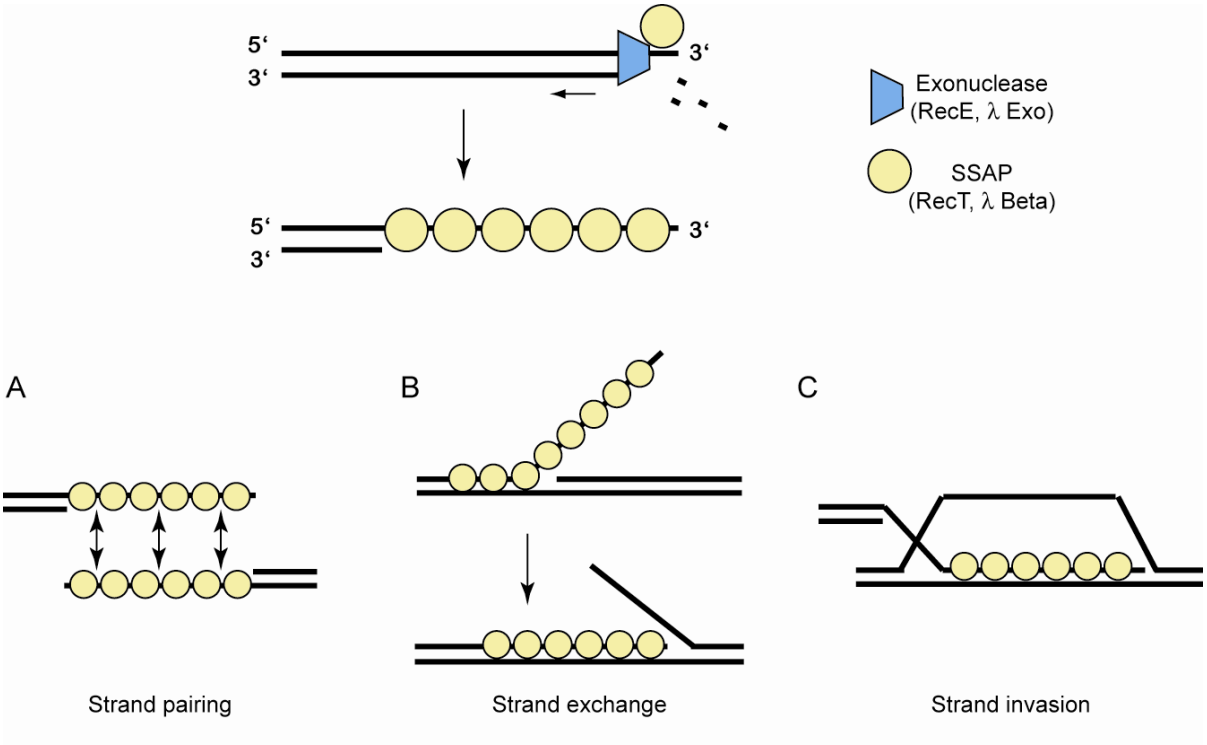


Figure 6. SSAPs can catalyze recombination by three basic mechanisms: (A) strand pairing, (B) strand exchange, and (C) strand invasion. The partner exonuclease (RecE or Exo) degrades a dsDNA end 5′-3′ leaving behind a 3′ ssDNA tail. This is bound by the SSAP (RecT or Beta) and recombined with its homologous target sequence.

1.2.1 Single strand annealing protein families

The founding members of the SSAP superfamilies – λ Beta, Rac RecT, P22 Erf, and yeast Rad52 – have been extensively characterized genetically, biochemically, and structurally, leading to the general concept that these proteins are functional analogues and 'structural homologues' [162]. These 'recombinases' form ring structures, bind ssDNA and dsDNA, and catalyze pairing, strand exchange, and strand invasion [162,174,204,224]. Although no sequence similarity was initially observed between any of the founding members, they were shown to fall into three evolutionarily defined superfamilies [85]. The 'Red β/RecT superfamily' is comprised of the bacteriophage λ Beta (Red β) and the *E. coli* Rac prophage RecT proteins. RecT and Beta have no apparent sequence similarity but function analogously such that RecET can substitute for Exo/Beta for phage λ recombination [66]. PSI-BLAST analysis with Beta homologues from numerous other lambdoid phages retrieves the RecT protein and its homologues. Sequence analyses further revealed several conserved residues as well as secondary structure predictions that correlate well with some of their biochemical properties, such as Mg^{2+}-dependent ssDNA-pairing and dsDNA binding activities [96,145]. Further, λ Beta homologues are present in numerous diverse bacteria and phages, while RecT-like proteins appear predominantly in low G+C% Gram-positive bacteria and phages. Two proteins found in this superfamily – *E. coli* EHAP1 and *Borrelia hermsii* PF161 – have an unusual domain structure; the N-terminal domain is similar to the Beta/RecT family, while the C-terminus is similar to the Erf family.

The bacteriophage P22 Erf protein has also been described as a SSAP and defines another superfamily [85,178]. Conserved motifs have been identified in these proteins, and much like the

Beta/RecT family, they seem to have originated in bacteriophages and subsequently appeared in bacterial genomes as prophages. P22 Erf can also substitute functionally for λ Beta [175].

The third small superfamily of both eukaryotic and bacterial SSAPs was identified by database searches with eukaryotic Rad52 proteins. Rad52 from yeast and humans has been shown to act as a SSAP in conjunction with the RecA ortholog Rad51 [17]. Sequence alignments and structural predictions detect a conservation of two large motifs and other structural elements (including two putative helix-hairpin-helix folds) in both eukaryotic and bacterial Rad52s, indicating that they all belong to a single superfamily [85]. Although these proteins have been characterized biochemically, the following sections will focus on the bacteriophage systems of phage λ, *E. coli* Rac prophage, and P22.

The genes adjacent to the SSAPs are commonly predicted to be DNA recombination or repair proteins [85]. These include single-strand-binding protein (SSB), Holliday junction resolvases, and nucleases, specifically exonucleases like λ Exo and RecE, which are found with λ Beta and RecT, respectively. Most of the exonucleases fall in two families, the type II restriction enzyme fold (*e.g.* λ Exo) and the type EndoVII fold. This suggests that SSAPs work in conjunction with their partner proteins in recombination and recombinational DNA repair. However, in some phages, the SSAPs and exonucleases are mixed, which is unexpected given the apparent specificity of the exonuclease-SSAP protein interaction observed with the λ Red and RecET systems [142]. For example, in several phages, a gene encoding a λ Exo-like protein is located next to a RecT-like gene. Additionally, SbcC-like genes are adjacent to both RecT- and λ Beta-like genes. In one unique case, a Beta-like gene was fused to a C-terminal fragment of the P22 Erf gene (*Borrelia hermsii* circular plasmid *pf161* gene). Also, the order of the genes within the operon differs between phages such that either the SSAP or its partner gene may be

transcribed first [43,85]. Collectively, the organization of these phage-encoded recombination genes reflects the modular structure that is characteristic of phage genomes.

1.2.2 The λ Red recombination proteins

The Red recombination system of bacteriophage λ was identified by the observation that bacteriophage λ could replicate in the absence of RecA [23]. Red$^-$ mutants (recombination-deficient) were found to map to genes encoding the Exo and Beta proteins [181,207], which were shown to be required for the RecA-independent recombination observed in λ [206,207]. Red-mediated recombination is stimulated by the presence of double-strand breaks that act as the substrates for Exo. Exo is an ATP-dependent dsDNA exonuclease that degrades DNA in the 5′ to 3′ direction at approximately 1000 bases per second [29,63,116,124] and leaves behind long 3′ ssDNA ends [79]. The enzyme requires a dsDNA end for activity and cannot degrade at nicks in DNA [28,29]. The structure of the active enzyme is a trimer that forms a toroid through which the dsDNA passes at one end and the resulting ssDNA substrate through the other [100].

The λ Beta protein is a SSAP that binds ssDNA substrates of lengths greater than 35 nucleotides [144] that protects ssDNA from nuclease attack prior to synapsis [92,114,129]. Beta promotes renaturation of complementary ssDNAs [96,129], strand exchange (displacement) [114], and strand invasion [193] all of which have been studied as recombination mechanisms of the single strand annealing pathway. Following pairing of ssDNAs, Beta binds tightly to the dsDNA complex [114]. Electron microscopic analyses show that Beta – like RecT and P22 Erf – forms circular structures in the absence of DNA which increase in size and monomer composition in the presence of ssDNA. Beta also forms helical filaments in the presence of dsDNA [162]. The data from structural studies suggest that ssDNA molecules are actually

wrapped around the Beta toroid, perhaps to prevent ssDNA from forming secondary structure and maintaining a conformation such that the bases are exposed for strand pairing. Beta also interacts with other proteins as determined by co-purification which precipitate λ Exo [129,182], host ribosomal protein S1 [129], and RNA polymerase subunit NusA [231]. Beta interacts specifically with Exo [129], functioning to modulate its activity as it degrades linear dsDNA substrates [225], and this interaction cannot be mimicked with other functionally analogous exonucleases [142]. Another attractive idea is that Beta also functions to interact with transcription and translation factors, perhaps to remove these complexes in front of the exonuclease [106].

The third protein that acts with the Red system in λ, Gam, binds the RecB subunit of the RecBCD nuclease in *E. coli*, preventing it from binding to dsDNA ends and thereby inhibiting all known enzymatic activities of this complex [39,93,122,133,138,176]. It has also been shown to interact genetically with the gene product of *sbcC*, though this is less-well characterized [102]. Although Gam is not required for recombination activities of Exo and Beta in phage λ [53], it increases recombination by limiting host nuclease attack on linear dsDNA substrates [38,106,142,240,241]. Alternatively, strains of *E. coli* that are Δ*recBC* Δ*sbcBC* or Δ*recD* (which are typically used for linear DNA transformation) show an increase in recombination of linear AESs, though not as high as observed using a Gam-expressing strain (20- to 800-fold increase) [135]. Numerous other bacteriophages are known to encode Gam functional analogues that inactivate or block host nucleases [195], and examples include the phage T4 protein gp2 [6,115,208] and phage Mu Gam protein [2]. These proteins bind dsDNA ends and protect injected linear DNA from degradation by RecBCD. In addition, the phage P22 Abc1 and Abc2 proteins work cooperatively to modulate RecBCD activity (discussed below). Therefore,

although the mechanisms of nuclease inhibition are different, the ultimate result is the protection of linear DNA ends from degradation by host nuclease.

The Red genes *exo* and *bet,* along with *gam*, are expressed from the P_L operon during early infection or upon induction of lysis of the prophage [38]. The Exo and Beta proteins are believed to play a role in phage λ infection during DNA replication by functioning to increase DNA synthesis [106], although this is still not well understood. Phage λ DNA molecules are replicated initially as circular molecules by theta replication, and this switches to rolling circle (sigma) replication and forms concatemers of linear DNA. Since initiation of DNA replication likely requires circular DNA (prior to concatemer formation), DNA synthesis could conceivably be increased through generation of additional circular genomes by Exo/Beta recombination [106]. Gam functions to inhibit the degradation of the linear concatemers of λ genomic DNA by the RecBCD nuclease [53]. Therefore the Red and Gam proteins are not essential for λ propagation but mutations in these genes result in fewer plaques [53]. The Red proteins also are involved in generalized transduction of λ, although at low levels compared to RecA [106]. Additionally, since the conditions that stimulate the lytic cycle of λ prophages may also cause DNA damage (such as ultraviolet light), the Red proteins could be important for repair of the resulting double-strand breaks [179]. Numerous studies have investigated the mechanism of recombination employed by phage λ [106], and it has been found that single strand annealing occurs in the absence of RecA, while strand invasion is favored in the presence of RecA [216].

1.2.3 The Rac prophage RecET recombination proteins

SSAPs were first described in *E. coli* as an alternative recombination pathway in a Δ*recBC* strain [12]. Analysis of mutations that suppress *recBC* revealed a class of mutations that

map to the *sbcA* gene (suppressor of *recBC*) and activate expression of the *recE* and *recT* genes of the cryptic Rac prophage in the *E. coli* genome [34,68,105]. The RecE (ExoVIII) and RecT (RecET) proteins catalyze recombination independently of RecA similar to λ Exo and Beta and have been shown to be functional analogues. Specifically, mutants of λ deleted for the Red recombination genes were able to recombine only in *E. coli* strains that expressed the Rac prophage *recE* and *recT* genes (*i.e. sbcA*-) [63,66]. Although the two systems function similarly, recombination does not proceed when the paired proteins are mixed heterologously (*e.g.* λ Exo and RecT), and only RecE binds RecT *in vitro*, indicating that there is a specific interaction between the cognate proteins required for recombination [142].

The RecE enzyme – like λ Exo – is a highly processive ATP-dependent exonuclease that degrades linear dsDNA 5′ to 3′, cannot act at nicks or gaps, and has low but detectable activity on ssDNA [89,90,105]. RecE is a member of the RecB nuclease family of proteins: the C-terminus of RecE is similar to the nuclease domain in the C-terminus of RecB, and mutations in the conserved critical residues of RecE either abolish or decrease nuclease activity [31]. However, the N-terminal 587 amino acids (full-length RecE is 866 amino acids) are not required for its exonuclease activity or recombination [33,119,142]. The SSAP, RecT, acts to pair ssDNA substrates and promote strand exchange and invasion [68,69,145], and exhibits properties of homology-recognition with RecA [146]. It was also shown to bind dsDNA in the absence of magnesium, whereas ssDNA binding is only decreased slightly by the presence of magnesium [145]. RecT protein monomers form open and closed rings in the presence and absence of ssDNA as well as nucleoprotein filaments with RecE on dsDNA [224]. Finally, unlike phage λ, the Rac prophage does not encode a Gam-like protein.

1.2.4 The P22 Erf, Arf, and Abc recombination proteins

Much like λ, bacteriophage P22 encodes a homologous recombination system that functions through the single strand annealing pathway. However, unlike λ, recombination-mediated circularization of the linear genomic DNA upon entry into the host cell is required for DNA replication [237,238], and the phage proteins are therefore absolutely essential in *recA* strains of *Salmonella* [218]. Recombination-deficient mutants of P22 can also be complemented by the λ Exo and Beta proteins and vice versa [175,178]. The P22 recombination system is composed of Erf (essential recombination function), Arf (accessory recombination function), and Abc1 and Abc2 (anti-recBCD) proteins. Erf, the SSAP in this system, binds and protects ssDNA [131,173], promotes strand annealing [136], and forms ring structures [162,174]. It has also been shown to bind dsDNA under certain conditions [173], and in general appears to be biochemically equivalent to λ Beta and RecT. Arf is less well-characterized, but it is known to be required along with Erf for the recombination activity of P22, and is located adjacent to *erf* in the P_L operon [177,203].

The Abc proteins function to modulate RecBCD activity: they are not essential but phages lacking these are decreased in burst size [132]. Null mutations in *recB* of the host restore progeny levels to wild type [54], suggesting that they prevent the degradation activity of RecBCD much like λ Gam. It appears that Abc2 functions similarly to Gam but with distinct differences: Gam inhibits all activities of RecBCD [133], while Abc2 inhibits RecBCD recombination (dsDNA-exonuclease, ATPase, and helicase activities) but retains its 5′ ssDNA exonuclease activity [134]. It therefore appears that P22 uses Abc2 to modulate and exploit the ssDNA exonuclease activity of RecBCD to synthesize recombinogenic substrates for Erf. Through binding to the RecC subunit, this Abc2-modified RecBCD complex was shown to

interact with λ Beta and substitute for λ Exo in Red recombination [136]. Therefore the Abc2-RecBCD complex appears to have activity similar to λ Exo, RecE, and other 5′-3′ exonucleases. It is not clear yet what role the Arf and Abc1 proteins play, although Abc1 is not required for Abc2-RecBCD/λ Beta recombination of phage λ [136]. Further, it is unknown if one of the P22 recombination proteins or a host protein such as SSB functions to protect the 3′ ssDNA tails following degradation by Abc2-RecBCD. Finally, while the λ Red recombination proteins Exo and Beta can work independently of Gam, it is clear that the mechanism of recombination in P22 is different and requires its 'Gam analogue,' Abc2, for recombination. In fact, it appears functionally equivalent to both λ Exo and Gam by simultaneously inhibiting deleterious effects of RecBCD and taking advantage of its exonucleolytic activity in single strand annealing recombination.

1.2.5 SSAP mechanisms of recombination *in vivo*: single strand annealing versus strand exchange

Recombination by phage λ can proceed effectively in the absence of RecA [23,206,207], and both strand annealing and strand invasion activities have been shown in numerous *in vitro* reactions with only λ Beta or RecT proteins [69,96,114,129,145,193]. Yet, in different reports, the question as to which mechanism of recombination occurs *in vivo* has been contested [55,142,206,207]. Experiments investigating phage λ recombination have further implicated a role for RecA in some SSAP-mediated recombination such as strand invasion [64,96,129,178]. In studies where λ DNA replication is blocked, Red-mediated recombination is drastically reduced in the absence of RecA [215]. Thaler *et al.* showed that DNA replication of the λ genome was required to produce populations of dsDNA ends as substrates for the Red proteins

[222], which provided an explanation for λ Red recombination dependence on either DNA replication or RecA. Stahl *et al.* therefore carefully tested the two proposed mechanisms of Red-mediated recombination in λ: (1) strand invasion, and (2) strand annealing. They found that the strand annealing was the predominant type of recombination (with low levels of strand invasion) in the absence of RecA, whereas Red-mediated strand invasion occurred in the presence of RecA [216]. Strand annealing by the Red proteins was observed at a high frequency during λ DNA replication. However, λ Red dependence on DNA replication was eliminated by the introduction of dsDNA breaks on the λ genome, although this slowed strand annealing [216]. These data suggest that DNA undergoing replication is an optimal substrate for single strand annealing promoted by Red proteins, and likely other SSAPs.

1.3 RECOMBINEERING IN *ESCHERICHIA COLI*

Recent advances in *E. coli* genetics have illustrated the utility of bacteriophages through the development of a simple yet powerful technique called 'recombineering': genetic engineering in bacteria using phage recombination proteins [38,223]. Recombineering facilitates numerous types of mutagenesis in *E. coli* through expression of the potent recombination proteins of either the λ Red or Rac prophage systems. Single strand annealing recombination mediated by these proteins occurs with small lengths of homology (<50bp) and therefore allows simple synthesis of substrates for mutagenesis. This is reminiscent of genetic techniques that have long been available in yeast, in which the double-strand break repair system – that includes the SSAP Rad52p – promotes recombination between short regions of homology [152]. Recombineering in bacteria can be used to target chromosomes, plasmids, and phage genomes

and has been expanded for use in other Gram-negative bacteria such as *Salmonella, Shigella,* and *Vibrio* [42,185]. In addition, modifications of bacterial artificial chromosomes (BACs) by recomineering in *E. coli* has made a huge impact on functional genomic research [109,140,219,236]. This has simplified construction of mouse knockout constructs in BACs and high-throughput manipulation of genomic libraries by alleviating the time-consuming steps of traditional recombinant DNA cloning techniques [36,199]. Genetic engineering with ssDNA substrates has even been demonstrated in mammalian cells either expressing λ Beta or RecT [244] or in wild type cell lines [83,171]. Clearly this is a highly efficient system for genetics that is broadly applicable.

1.3.1 Recombineering systems: λ Red and RecET

E. coli recombineering systems have been successfully developed using both the λ Red/Gam proteins and the Rac prophage RecET proteins. This technique has far surpassed those previously available for targeted gene replacement by largely increasing the numbers of transformants that are recovered. Earlier methods used conventional AESs with large amounts of homology (>1 kbp) that were typically transformed into recombination-proficient *E. coli* such as Δ*recBC* Δ*sbcBC* or Δ*recD* strains [192], although this severely limited the strain background that could be utilized. In the first demonstration of recombineering, Murphy placed the λ *exo bet* genes in the chromosome of a Δ*recBCD* strain background and showed a large increase (up to three orders of magnitude) in gene replacement frequencies, which was dependent on inducible expression of Exo and Beta [135]. It was also shown that a strain expressing Exo, Beta and Gam worked just as well as a Δ*recBCD* strain expressing only Exo and Beta [135,137]; however,

expression of Exo and Beta in a wild type background was not sufficient to promote gene replacement without Gam or Δ*recBCD* in this particular study [135]. Murphy's 'hyper-rec' strain with the λ genes placed on the chromosome was more effective for recombination than strains containing the plasmid-encoded Exo/Beta, and therefore the decrease in copy number and level of protein expression in the chromosomally-encoded proteins was compensated [135]. One possible explanation for this was that perhaps the linear multimeric plasmids that undergo rolling circle replication compete with the linear AESs for Exo and Beta.

Shortly following this study, Zhang *et al.* produced a similar tool for gene replacement in *E. coli* using the RecET proteins in combination with λ Gam on a plasmid [242]. This system was developed following the observation that gene replacements were obtained with short homologies (42 bp) only in Δ*sbcA E. coli* strains, which express the RecET proteins from the Rac prophage. The need for an easily transferable system was solved by expressing the *recE* and *recT* genes from a plasmid. Further, the λ *gam* gene was incorporated in place of using Δ*recBC* strains.

Numerous technical advances were applied to these two methods, but ultimately the system developed with λ Red by Donald Court and colleagues was preferable and is now the most commonly used. A modified λ prophage was used to tightly control expression of *exo bet gam* for short induction times while preventing cell death from prolonged expression [240]. Including Gam in the recombination system eliminated the need for Δ*recBCD* strains and allowed high levels of recombineering in any strain background. These modifications eliminated problems with leaky expression that caused other undesirable recombination events and plasmid instability. While the λ prophage configuration is typically used, similar plasmid versions have been developed for use in *E. coli* and other bacterial systems [42], making the system more

mobile. The P22 system was also tested for its ability to promote recombineering in these assays but was found to be less efficient than the λ Red system [135], although this was not tested extensively.

1.3.2 The recombineering strategy for mutagenesis

Recombineering in *E. coli* has been successful for making several kinds of mutants: targeted gene replacements, point mutations, deletions, and small insertions [38,244]. The system can also be used for BAC modification, gene specific random mutagenesis, and *in vivo* cloning by gap repair [38,140,143,201,243]. Most applications of this system have been described in detailed protocols [7,38,42,201], though more are likely to appear in the future. Some of the more commonly used techniques such as targeted gene replacement and point mutagenesis will be discussed here in more detail.

Several expression strategies were tested for optimal recombination activity. In one setup, RecT was placed under a constitutive promoter and RecE under an inducible promoter [242]. Stronger promoters (P_{tac}) increased expression five-fold but actually decreased recombination activity two-fold [137]. Observations such as this indicated that there is likely an optimal level of expression of the pair of recombination proteins, and it is suggested that a 5:1 ratio of Beta to Exo results in the highest level of recombination (K. Murphy, personal communication). Others have placed the *exo bet* genes under inducible control while keeping *gam* constitutively expressed [141]. However, the ideal configuration for the λ Red/Gam system was developed using a modified prophage that carefully controls expression through their native promoter for maximal recombination activity and minimal cell death [240]. This was accomplished by removing the lytic genes and using a temperature-sensitive allele of the λ cI repressor (*cI857*)

such that expression of the P_L operon (including the λ *exo bet gam* genes) for less than 60 minutes is tolerable. The strain is grown at 32°C then shifted to 42°C for 15 minutes to induce expression of the Red proteins, after which electrocompetent cells are prepared [201]. Although the protocols differ for each type of mutagenesis, the strain background is typically the *cI857* λ defective prophage version, unless a plasmid encoding the λ Red genes is being used.

1.3.2.1 Recombineering with dsDNA substrates

Targeted gene replacement by recombineering eliminates the need for special recombination-proficient strains of *E. coli* and yields large numbers of colonies (>10^4) following transformation with an AES. Even the synthesis of the AES was made simpler by eliminating the need for cloning. Since the SSAPs can perform recombination with short substrates (>35 nt) [144], AESs can be made using PCR-generated substrates with short regions of homology flanking the antibiotic cassette (Figure 7) [242]. The distance between the homologies does not appear to affect recombination frequencies [242], while extending the length of homology from 20 bp to 40 bp increases recombination by four orders of magnitude [240]. Extending the homology proportionally increases gene replacement frequencies mediated by RecET or λ Exo/Beta up through 1500 bp [142], though the difference between 40 bp and 1000 bp only increased Red recombineering frequencies 10-fold [240]. However, since small regions of homology are sufficient, substrates for targeted gene replacement typically include 50 bp of homology flanking a variety of antibiotic resistance genes [7]. Saturating amounts of the AES are reached at 100 ng, so this quantity of DNA is used in standard transformations [240]. Counter-selection with *sacB* has also been used to generate unmarked deletions of genes following recombineering [137]. Further, this method for gene replacement can be used on either the chromosome or on plasmids [240].

Figure 7. Strategy for targeted gene replacement by recombineering.

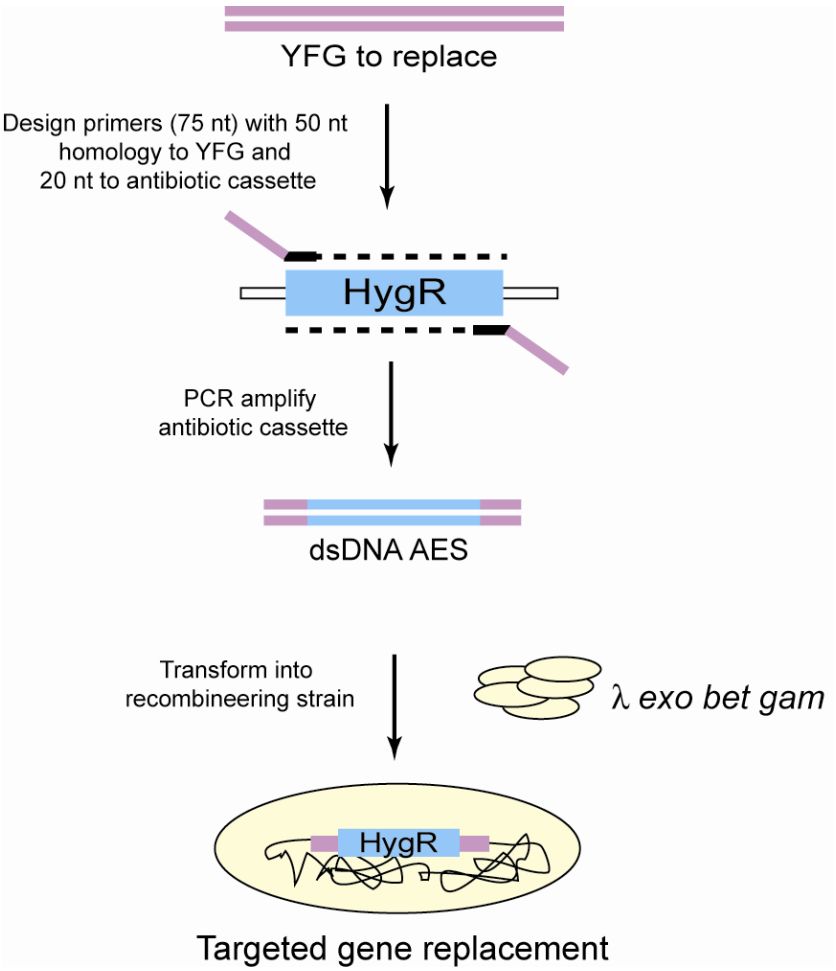


Figure 7. (adapted from Court *et al.* [38]). Primers (75 nt) are designed such that 50 nt at the 5′ end are homologous to the target gene (your favorite gene; YFG) and 20-25 nt at the 3′ anneal to an antibiotic resistance gene. PCR performed with these primers yields a dsDNA AES product with 50bp homology flanking the antibiotic resistance gene. Transformation of this AES into recombineering cells induced for expression of λ *exo bet gam* yields targeted gene replacement mutants by homologous recombination.

The dependence on host RecA, as well as on the individual Red proteins, was examined by measuring targeted gene replacement frequencies in strains missing any one of these [142,240]. A 10-fold drop was observed in Δ*recA* strains, indicating only a modest role for RecA in λ Red-mediated recombination. Deletion of any one of the Red- or Gam-encoding genes results in zero transformants as compared to 4,000 in a strain with all three. The dependence on Gam or the requirement for a Δ*recBCD* strain was observed in other studies [135,142]. However, a recent study showed that Gam is not required for recombineering of dsDNA substrates, although it increases recombineering frequencies ~10-fold [43].

1.3.2.2 Recombineering with ssDNA substrates

Point mutagenesis by recombineering was an important development that requires the simplest of manipulations. The ability to construct single nucleotide changes has numerous applications, including the study of specific amino acid effects on protein function and structure. This has been accomplished in *E. coli* on the chromosome, plasmids, and BACs using short ssDNA substrates [52,219]. Since SSAPs – like λ Beta – can bind and recombine short segments of ssDNA, point mutations can be made with synthetic ssDNA substrates. Oligonucleotides are synthesized containing a point mutation and are transformed into electrocompetent recombineering cells induced for λ Red/Gam expression. Point mutations by ssDNA recombineering are incorporated at a sufficiently high frequency to eliminate the need for selection (as high as 6% of the total survivors of electroporation in some experiments [52]). However, this number seems to be achieved only rarely, and more typical frequencies are 0.1% - 0.5%. Alternative strategies for introduction of point mutations have been developed that include a selection step [141]. The target gene is marked first with an antibiotic resistance gene and *sacB*

by targeted gene replacement. This allele is subsequently targeted by recombineering using a ssDNA that deletes the selection markers and simultaneously incorporates a point mutation; negative selection for cells that have lost *sacB*, which are able to grow on sucrose, identifies the mutant recombinant. Other strategies for identifying point mutants exist, such as the use of specialized PCR screens (mismatch amplification mutation assay; discussed in section 3.4.6) and inactivation of mismatch repair.

Recombineering can also be used to make deletions and small insertions. Deletion of the *galK* gene using a ssDNA substrate was shown to be as efficient as making a point mutation [52]. These ssDNA substrates can also be used to delete larger regions; this is particularly useful for the removal of antibiotic resistance and *sacB* genes in mutant strains [223]. Small insertions can be made, although the frequency of recombination decreases as the length of the insertion increases (tested up to 60 nt) [244]. The recombineering technology can likely be used for numerous other applications, and more developments will probably arise in the future.

The frequency of incorporation of point mutations is highly correlated with the activity of the *E. coli* methyl-directed mismatch repair (MMR) system [37]. Since the MMR proteins function to correct errors during DNA replication, repair of the recombineered point mutation back to wild type can occur at high frequencies by MMR. Elimination of MMR by mutation of *mutH, mutL, mutS, uvrD* results in an increase in ssDNA recombineering (25- to 60-fold). Recombineering frequencies with ssDNA were found to correlate with the pattern of MMR activity, such that certain mismatches are more frequently corrected than others. Ultimately, Δ*mutS* strains are recommended for increasing ssDNA recombineering frequencies for point mutagenesis in up to 25% of viable cells following transformation [37].

The dependence of ssDNA recombineering on the length of the ssDNA substrate has also been tested. In one study, maximal numbers of point mutants were obtained in strains expressing Beta with a 70 nt substrate; shortening these to 60, 50, or 40 nt resulted in a large drop in recombination frequency (approximately four orders of magnitude), and lengths of 20 nt did not recombine [52]. The 10-fold decrease in recombination observed with oligonucleotides shortened from 40 nt to 30 nt [52] likely reflects the length requirement for Beta binding to ssDNA (36 nt) [144]. RecT was also found to recombine longer ssDNAs substrates more efficiently (>30 nt) [244]. In addition, homology was required on both sides of the point mutation, and placement of the point mutation at either the 3′ or 5′ end of the ssDNA substrate did not produce recombinants [244]. However, shifting the point mutation toward the 3′ end such that more homology was present on the 5′ side was more successful than the opposite scenario. These data suggest a requirement for binding of the SSAP on both sides of the substrate. It is also noteworthy that annealing two complementary oligonucleotides did increase recombineering frequencies slightly in some assays compared to using either oligonucleotide independently [244].

A variety of mutant host strains were tested by Zhang *et al.* to determine the contribution of host recombination proteins, and it was found that strains with Δ*sbcBC* mutations were more deficient for ssDNA recombineering than wild type [244]. In contrast to targeted gene replacements with dsDNA substrates, recombineering of a ssDNA substrate was not at all dependent on RecA [52,244].

It has been demonstrated that only λ Beta (or RecT) is necessary and sufficient for ssDNA recombineering [52,241,244]. However, another study with RecET and λ Exo/Beta showed a slight increase in frequency when the cognate exonuclease was included with the

SSAP [244]. This is further evidence of a specific protein-protein interaction between these pairs of proteins [142]. Since SSAPs are found in a plethora of prokaryotes and eukaryotes, it was postulated that this type of recombination could be extended into other systems, and indeed it was shown that both λ Beta and RecT function in mouse ES cells to promote ssDNA recombineering [244]. The P22 Erf protein was also shown to function in ssDNA recombineering similar to λ Beta and RecT, although the P22 system was not found to support dsDNA recombineering [244]. In one study, deletion of *gam* resulted in approximately a five-fold decrease in ssDNA recombination frequency [52]. It is not clear why Gam is required for maximal recombination since Beta binds and protects ssDNA from nuclease attack. Yu *et al*. hypothesized that perhaps the ssDNA nuclease activity of RecBCD still has a slight negative effect. This observation was contradicted by another study in which no difference in ssDNA recombineering was found in the presence or absence of RecBC or λ Gam [244].

A strand bias was observed in correlation with ssDNA recombineering frequencies. Using oligonucleotides that anneal to both strands of the chromosome at six different loci, it was found that the oligonucleotide that annealed to the template for lagging strand (discontinuous) DNA replication (referred to as the 'lagging strand') was most efficient [52]. The biases toward ssDNAs targeting the lagging versus the leading strand ranged from 2- to 50-fold. This supports the hypothesis that the direction of DNA replication at the target locus directly influences the recombination frequency of ssDNAs, since the lagging strand likely has more single-stranded regions exposed to which a ssDNA substrate (bound by Beta) could anneal and recombine [38]. However, other cellular processes such as transcription, MMR, or other DNA repair systems that function with strand-specificity could conceivably generate exposed regions of ssDNA for pairing predominantly one strand at a particular locus, resulting in a strand bias. Numerous

reports that examined recombination with ssDNA substrates in yeast and mammalian cells present data that transcription plays a large role in the strand biases [83,117]. Therefore, Li and colleagues examined the effects of these different factors on ssDNA recombination [113]. They conclude that, in *E. coli*, MMR and DNA replication are the major contributors to the observed strand biases, with little to no influence from other cellular processes such as transcription.

Therefore, the current model for ssDNA recombination in *E. coli*, the 'annealing-integration' model, suggests that the ssDNA anneals to the lagging strand and DNA polymerase and ligase complete the reaction to join this ssDNA to the template (Figure 8A). Further, sequence-specific effects can be dominant to the role of DNA replication for mutations that are corrected by the MMR system. This model could also be extended to examine recombination of dsDNA substrates, in which the resected, SSAP-bound 3′ ends could also anneal to the lagging strand during DNA replication [38]. Previously, it was thought that dsDNA substrates were recombined either by strand annealing or strand invasion [106], but these mechanisms imply an indirect role for DNA replication to provide exposed ssDNA surfaces for recombination. Alternatively, while recombineering of dsDNA substrates is likely different than that which occurs with λ phage DNA recombination, a direct role of DNA replication would connect observations made of the two processes.

One model that has more experimental support is called the 'replisome invasion and template switch' mechanism (Figure 8B) [180]. This suggests that the SSAP-bound 3′ ssDNA end that is annealed to the lagging strand actually becomes a template for continuous (leading) strand synthesis. Replication continues through this substrate, and the lagging strand portion of the fork is released. However, this leaves several subsequent details unresolved, such as the fate of the unreplicated lagging strand half of the fork. A more likely model suggests that replication

does not continue through the substrate, but terminates at the dsDNA junction, and is completed by ligation (not shown). Following this, recombination of the second resected end results in replacement of the wild type template with the dsDNA substrate (K. Murphy, personal communication). These models are both currently being further tested.

Figure 8. Models for the mechanism of ssDNA and dsDNA recombineering.

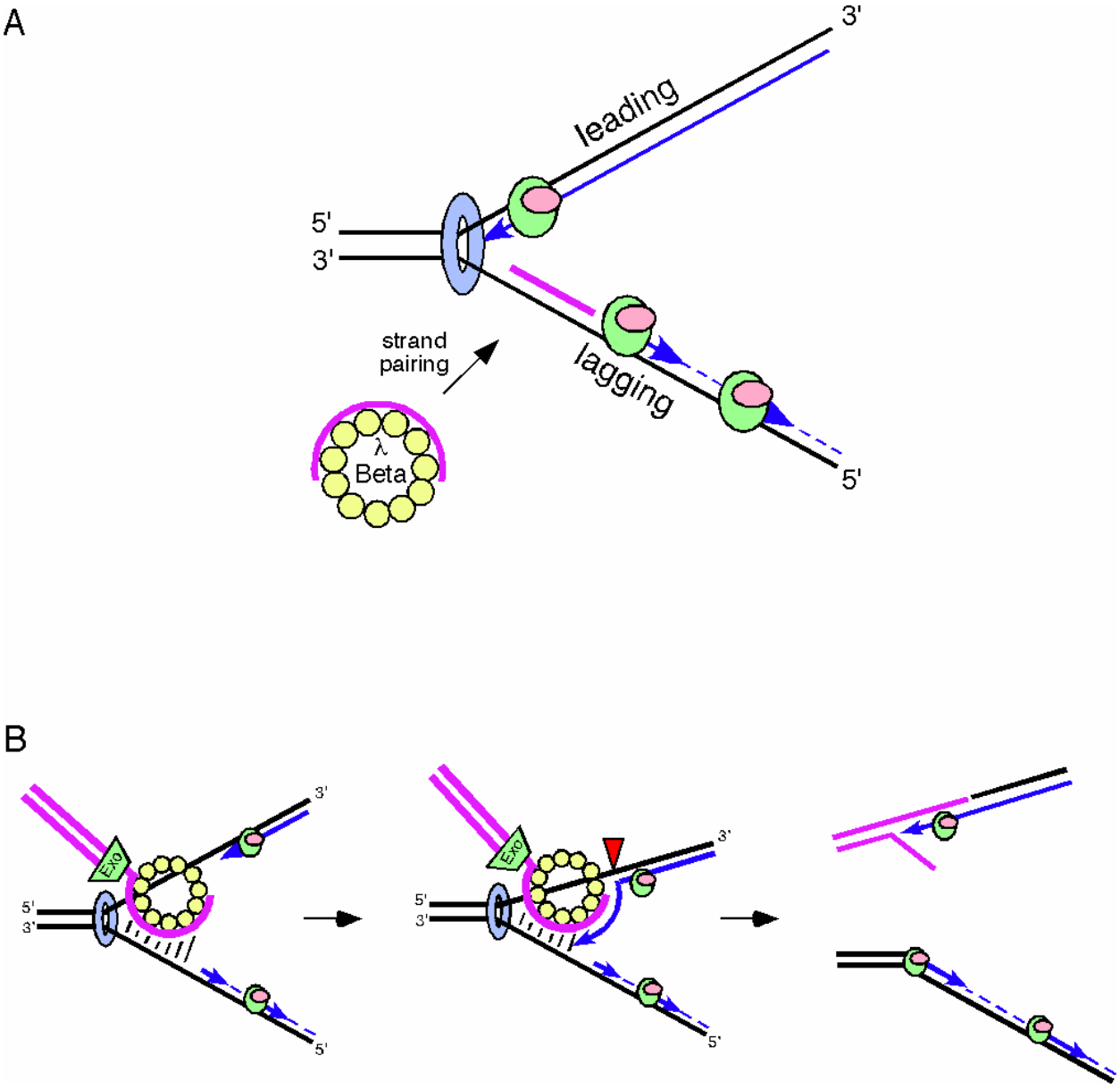


Figure 8. Models for how recombineering substrates might be incorporated during DNA replication. (A) During ssDNA recombineering, the SSAP (*e.g.*Beta) forms a toroid around which the ssDNA substrate is wrapped, and Beta promotes strand pairing with the chromosome. This occurs preferentially with substrates that anneal to the lagging strand where an exposed ssDNA template may be more available. (B) The 'replisome invasion and template switch' model for dsDNA recombineering. dsDNA substrates are degraded by the 5'-3' exonuclease (*e.g.* Exo), leaving behind a 3' ssDNA tail bound by Beta. This anneals with the lagging strand, and is positioned in line with the replicating DNA fork. This becomes a template for leading strand synthesis, and the original chromosomal template is cleaved. This results in displacement of the lagging strand, and continuous replication proceeds through the dsDNA substrate. Presumably a similar reaction occurs at the second resected site. (adapted from Court *et al.* [38] and Poteete [180])

1.4 SPECIFIC AIMS OF THIS STUDY

The development of a simple and efficient system for genetics would greatly benefit the mycobacterial research community. Numerous factors inherent to mycobacterial cell growth and cell wall structure prevent simple handling and manipulation of the mycobacteria. However, the relatively high levels of illegitimate recombination compared to homologous recombination in *M. tuberculosis* and other slow-growers is the primary limiting factor to the application of conventional genetic techniques in these bacteria. The current methods for targeted gene replacement are designed to circumvent illegitimate recombination by modifying the AES or its delivery into the host cell. It is striking, however, that none of these have focused on increasing the levels of homologous recombination within the bacterial cell. The historical success of adapting mycobacteriophages and their proteins for manipulation of their mycobacterial hosts has led to the hypothesis that mycobacteriophage-encoded recombination proteins could be introduced into the mycobacterial cell to improve the efficiency of homologous recombination and thereby promote allelic exchange for mutagenesis purposes. The success of the λ Red recombination system for recombineering in *E. coli* further supported this notion and provided a basis for initial experimental design. Therefore the focus of my thesis research has been to utilize mycobacteriophage-encoded recombination proteins to develop a recombineering system for the mycobacteria.

1.4.1 Specific Aim 1: Bioinformatic and biochemical analysis of mycobacteriophage Che9c-encoded RecET homologues.

Mycobacteriophage-encoded homologues of the *E. coli* Rac prophage RecET proteins are rare in mycobacteriophages; only Che9c was found to encode homologues of both. *In vitro* biochemical analysis of Che9c gp60 and gp61 demonstrates that they possess exonuclease activity and DNA binding activities, respectively, similar to RecET. These data are presented in Chapter 2, and some of the experiments have been published [227].

1.4.2 Specific Aim 2: Development of a mycobacterial recombineering system using mycobacteriophage Che9c-encoded recombination proteins.

Che9c gp60 and gp61 have biochemical properties reminiscent of a coordinated recombination system that functions via the single strand annealing pathway. Expression of these proteins in mycobacterial strains yields a substantial increase in homologous recombination. This has provided an efficient genetic tool that has been successfully used to construct gene replacement mutants and point mutants in the genomes of both *M. smegmatis* and *M. tuberculosis*, and likely is applicable to other mycobacterial species. Chapter 3 describes the development of the mycobacterial recombineering system and the various technical applications, the majority of which have been published [227-229].

1.4.3 Specific Aim 3: Identification of additional mycobacteriophage-encoded recombination systems.

Sequencing of more than 50 mycobacteriophage genomes has revealed several additional gene candidates that may encode functional recombination proteins; these are present in the genomes of phages Giles, Halo, Wildcat, and also prophages in the genome of *M. avium* and *Mycobacterium abscessus*. *In vivo* analysis of several of the putative SSAPs, as well as λ Beta and RecT, demonstrate that the Che9c gp61 functions most efficiently in mycobacteria. Mycobacteriophage TM4 also appears to encode a recombination system, although the genes responsible have not thus far been identified by bioinformatic analysis. Experimental analysis of TM4 cosmid recombination sheds some light on the mechanism of TM4 recombination *in vivo*. These experiments and the implications of the results are discussed in Chapter 4.

2.0 MYCOBACTERIOPHAGE CHE9C ENCODES RecE AND RecT HOMOLOGUES

2.1 INTRODUCTION

Bacteriophages are an extremely diverse group of organisms, and at an estimated 10^{31} total phage particles, they are more abundant than any other life form in the biosphere [77]. Phages can be found in a variety of environments along with their bacterial hosts, and interactions between phages and bacterial populations foster copious amounts of genetic exchange. This contributes to a large pool of shared genetic elements [76] and has a significant impact on the evolution of bacteria, particularly on mechanisms of pathogenicity and acquisition of virulence genes [233]. Phages are often grouped based on their morphology, host-range, and other types of limited characteristics. However, it has become apparent that relationships among phages are better represented and understood through examination of their gene similarity and organization and by grouping them in ways that account for both their high level of diversity and the independent origin of their genes [108].

Although the number of well-characterized phages is a miniscule fraction of the total population, more than 500 phage genomes have been sequenced to date [73]. A significant proportion of these include the group of phages that infect the mycobacteria: the mycobacteriophages. More than 50 mycobacteriophage genomes have been sequenced [73,126,165,170] (and G. Hatfull, unpublished data), revealing a mosaic architecture reminiscent

of that originally observed in the lamboid phages and in other phages [165]. In this way, when comparing phages, similar genes are often staggered amongst genes have been acquired in a different way, and are organized in a modular organization. The unique combination of genes and gene clusters in this manner – with little to no sequence homology at the gene boundaries – is evidence that illegitimate recombination plays an important role in genetic exchange of functional genetic elements [126]. An alternative hypothesis has been suggested in a recent study that homeologous recombination – recombination between sequences that are related but are divergent – contributed to genetic mosaicism in phage λ [123]. Further, Martinsohn *et al.* suggest that the λ Red recombination system contributes to this recombination substantially more than the host rec proteins. However, the contribution of this particular type of recombination may not be the common contributing factor in other phages, and the observations made in this article could be limited to a small number of phages. Since the presence and/or activity of these types of recombination systems has not been carefully examined in many phages, the effect and prevalence of homeologous recomibination is unclear.

Bioinformatic analysis of mycobacteriophage genomes indicates that the genes encoding phage structural and assembly proteins are typically organized in similarly ordered operons, and therefore their function can often be inferred from previously characterized genes [165]. However, approximately half of the mycobacteriophage ORFs do not have detectable similarity to known genes from either phages or other organisms, and their function is unknown [71]. Of the mycobacteriophage genes that do have homologues, 90% of these are found in other mycobacteriophages, indicating that these organisms exchange DNA more frequently amongst themselves than with their bacterial hosts or other phages. A large proportion of the genes that have detectable similarity to known genes are found in multiple mycobacteriophages, while a

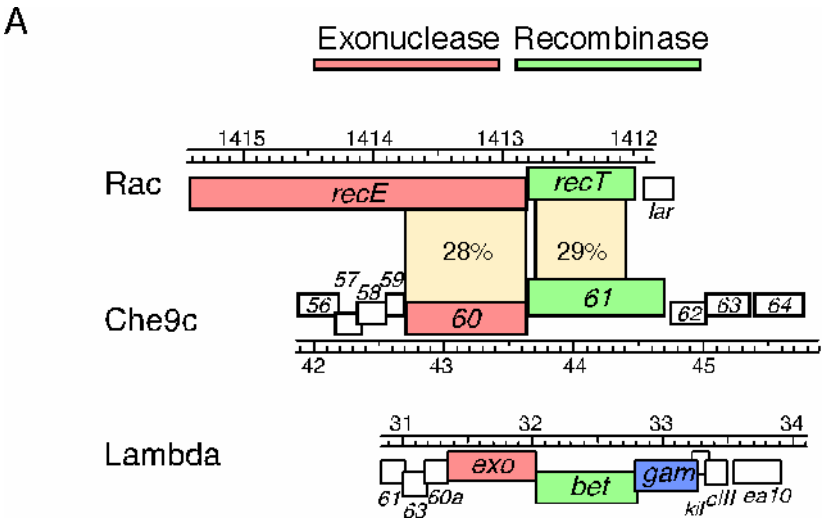
small number are homologues of genes from other organisms, including bacteria and other phages. Thus, sequencing of this relatively small number of phages has revealed a largely untapped reservoir of genetic information, suggesting that characterization of phage genomes is important not only to gain evolutionary perspectives, but also to explore and exploit the diversity of their gene pool.

Phage-encoded SSAP genes are examples of the architectural modularity found in bacteriophage genomes [85]. Most SSAP genes in the phages of the λ Beta/RecT superfamily are situated adjacent to DNA recombination or repair genes, although the pairing and operon organization of these differ in each phage. Identification of *E. coli* Rac prophage RecET homologues in mycobacteriophages illustrates not only a mosaic architecture but also the relatively rare occurrence of these types of genes in mycobacteriophage genomes. Initially, out of 14 sequenced mycobacteriophage genomes [73], only Che9c was found to encode both RecE-like and RecT-like gene products [165]. Further sequencing of mycobacteriophage genomes revealed additional ORFs with homology to proteins from known recombination systems, and these will be discussed in Chapter Four. Discovery of the Che9c recombination proteins suggested that these might be utilized to develop recombineering in the mycobacteria. Therefore, biochemical analysis was undertaken to examine the properties of the Che9c proteins to see if they function similarly to the RecET proteins.

2.2 BIOINFORMATIC ANALYSES OF MYCOBACTERIOPHAGES REVEALS A PUTATIVE RECOMBINATION SYSTEM

Through BLAST analyses [4], it was observed that identifiable recombination systems are rare in the mycobacteriophages, and only one phage encodes proteins that are distantly related to RecE and RecT of the *E. coli* Rac prophage (Figure 9). Che9c gp60 shares 28% identity with the C-terminal region of RecE. This encompasses a nuclease domain belonging to the RecB family, while the N-terminus of RecE is not necessary for its exonuclease activity [31]. The N-terminal two-thirds of gp61 (residues 28-237; Figure 9A) have 29% identity to RecT, whereas the C-terminal third of gp61 (residues 238-353) only has detectable similarity to the corresponding region of a predicted *M. avium* RecT protein (discussed in Chapter Four) and no other known proteins. A multiple sequence alignment performed with Che9c gp61 and the proteins identified by Iyer *et al.* as members of the λ Beta/RecT superfamily shows conservation of a core domain (200 amino acids) and a similar predicted secondary structure (Figure 9B), indicating that gp61 is indeed a member of this superfamily of SSAPs [85]. Much like the Rac prophage RecET system, no Gam homologues have thus far been identified in any of the sequenced mycobacteriophages.

Figure 9. Che9c gp60 and gp61 are RecET homologues.



B

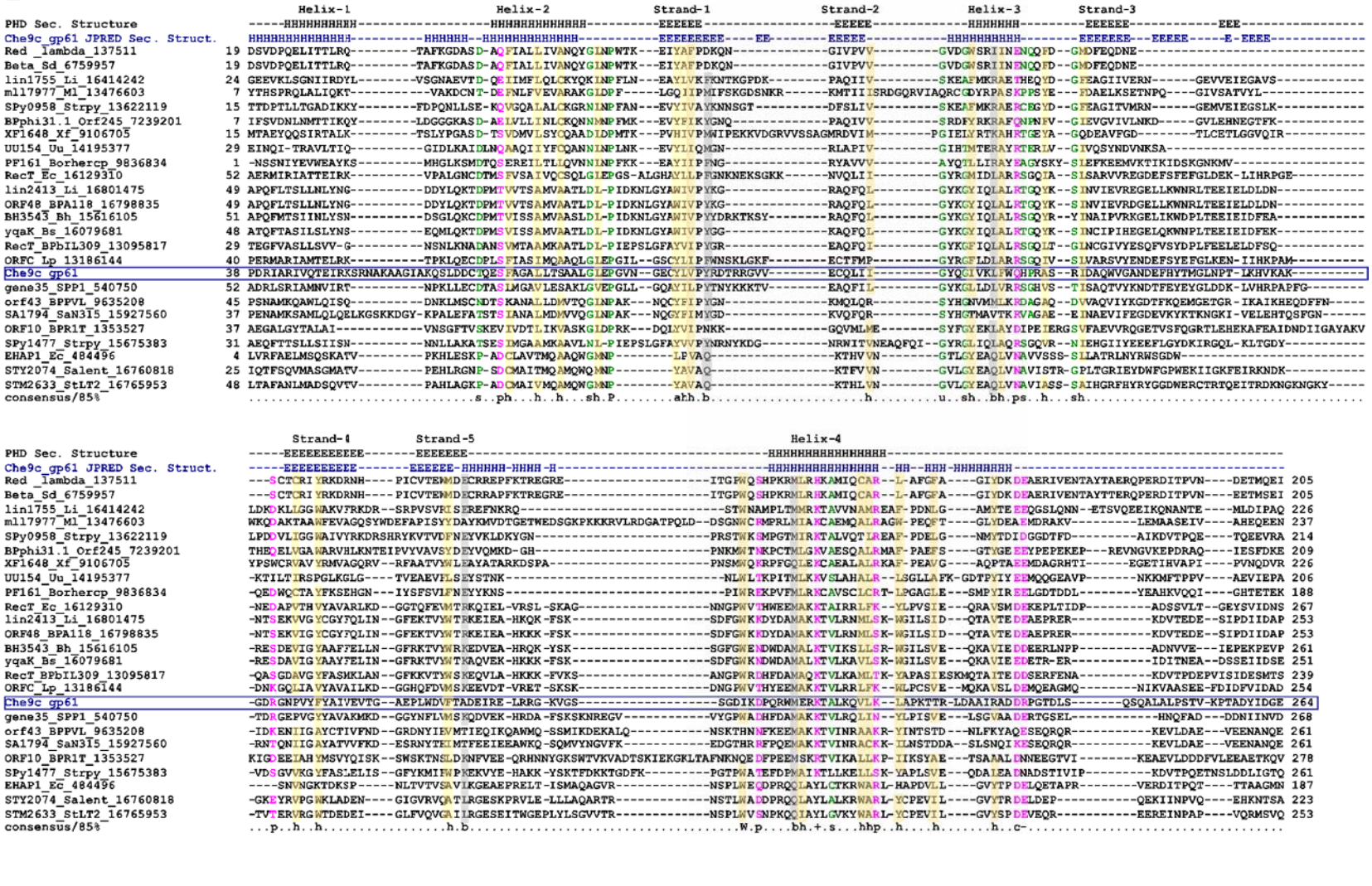

Helix-1
Helix-2
Strand-1
Strand-2
Helix-3
Strand-3
Strand-4
Strand-5
Helix-4
PHD Sec. Structure
Che9c_gp61 JPRED Sec. Struct.
Red_lambda_137511
Beta_Sd_6759957
lin1755_Li_16414242
ml17977_Ml_13476603
SPy0958_Strpy_13622119
BPphi31.1_Orf245_7239201
XF1648_Xf_9106705
UU154_Uu_14195377
PF161_Borhercp_9836834
RecT_Ec_16129310
lin2413_Li_16801475
ORF48_BPA118_16798835
BH3543_Bh_15616105
yqaK_Bs_16079681
RecT_BPbIL309_13095817
ORFC_Lp_13186144
Che9c_gp61
gene35_SPP1_540750
orf43_BPPVL_9635208
SA1794_SaN3I5_15927560
ORF10_BPR1T_1353527
SPy1477_Strpy_15675383
EHAP1_Ec_484496
STY2074_Salent_16760818
STM2633_StLT2_16765953
consensus/85%

Figure 9. (A) Che9c gp60 is a RecE homologue, while Che9c gp61 is a RecT homologue. Exonucleases are indicated in red, and SSAPs (recombinases) are indicated in green. *E. coli* Rac prophage genes and Che9c genes are transcribed from left to right, while the λ genes are transcribed right to left. (B) Multiple sequence alignments were performed with all protein sequences used by Iyer *et al.* [85], and conserved regions are shown. The T-coffee program was used to align Che9c gp61 (outlined in blue) with the λ Beta/RecT protein family members [148], and this was manually incorporated into the alignment made by Iyer *et al.* Secondary structure predictions (using JPred) for gp61 were also conserved (shown in blue at the top) [40]. Similar residues are highlighted that were found by Iyer *et al.* to be conserved greater than 85%: h, hydrophobic; l, aliphatic; a, aromatic; o, alcohol; c, charged; +. basic; -, acidic; p, polar; b, big; s, small; u, tiny.

2.3 PURIFICATION OF CHE9C GP60 AND GP61 PROTEINS

To determine if the Che9c proteins function similarly to their RecET homologues, gp60 and gp61 proteins containing C-terminal 6x-histadine tags were over-expressed and purified from *E. coli* lysates by nickel-affinity chromatography. SDS-PAGE analysis indicated that purified samples of recombinant gp61 were nearly homogeneous, while recombinant gp60 samples retained small amounts of contaminating host proteins (Figure 10A). Therefore, a mock-purification was performed with *E. coli* extracts from a strain containing an empty vector, and it was observed that these samples contained similar host proteins to the gp60 preparation (Figure 10B). This mock-purified protein sample was used for biochemical assays alongside gp60 as a negative control.

Figure 10. SDS-PAGE analysis of purified Che9c gp60 and gp61 protein samples.

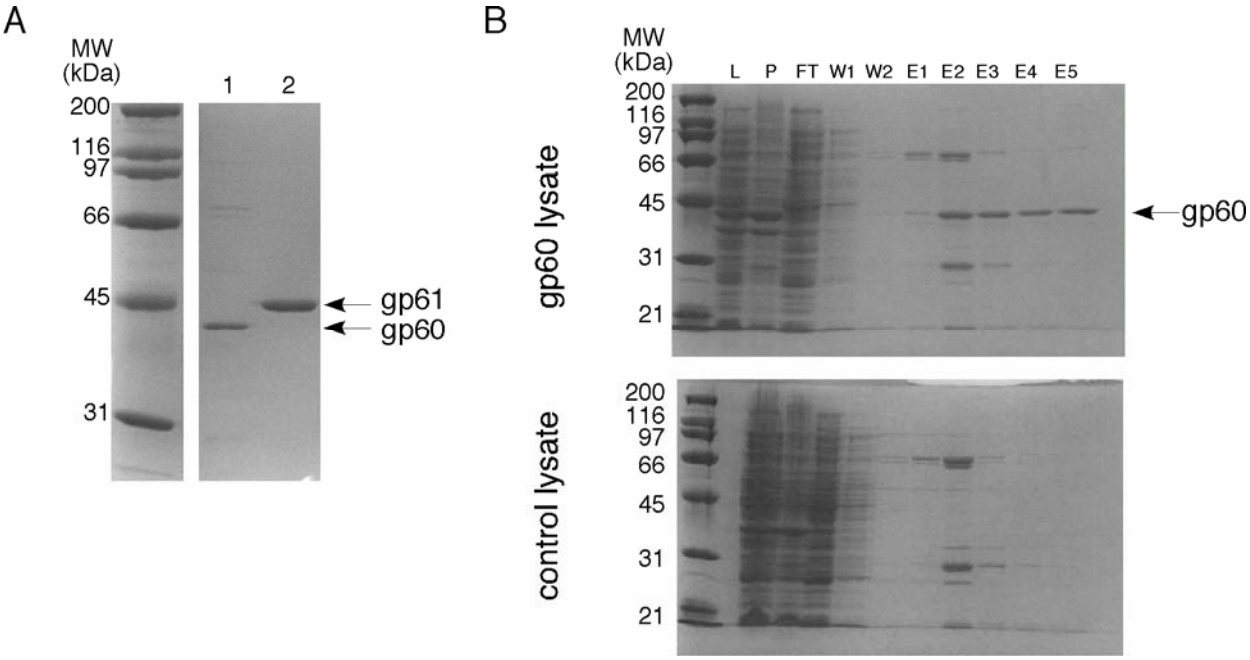


Figure 10. Recombinant gp60 and gp61 were over-expressed and purified from *E. coli* and samples analyzed by SDS-PAGE. Molecular weight (MW) in kDa is indicated by the standard protein ladder. (A) Approximately 0.5 μg of protein samples were loaded on this gel. Che9c gp60 (36 kDa) was purified to a concentration of 0.1 mg/ml (2.5 μM), although contaminating proteins were observed. Che9c gp61 (40 kDa) was purified to a concentration of 3.87 mg/ml (96 μM). (B) Eluates from mock-purified *E. coli* lysate from a strain containing an empty vector (control lysate) contains similar proteins that contaminate the preparation from *E. coli* lysates of strains expressing gp60 (gp60 lysate). The mock-purified sample was used as a control for gp60 and was stored at a concentration of 4 μg/ml. Protein samples from the last two elutions from each lysate were dialyzed and stored. L, lysate; P, pellet; FT, flow-through; W, wash; E, elution.

2.4 CHE9C GP60 IS AN EXONUCLEASE

Phage λ Exo and *E. coli* RecE are highly processive enzymes that degrade linear dsDNA in a 5′ to 3′ direction [89,116]. To determine if gp60 has exonuclease activity, three *in vitro* assays were developed [227]. First, gp60 was observed to degrade short radiolabeled dsDNA substrates (100 bp) similarly to λ Exo, while no degradation was seen in negative control reactions (Figure 11A). Notably, it was observed that serial dilutions of gp60 did not yield the expected step-wise decrease in activity, but rather even a 2-fold dilution resulted in very little degradation activity, which may be due to protein inactivation in dilution buffers. Because the observed activity of gp60 could conceivably also be attributed to a contaminating phosphatase that would remove the radioactive phosphate, a similar assay with linearized plasmid dsDNA substrates was used and visualized by agarose gel electrophoresis. Incubation of gp60 with this dsDNA substrate also resulted in degradation, while negative control reactions with mock-purified protein did not (Figure 11B). Finally, the observed exonuclease activity was shown to be limited to substrates with dsDNA ends, since neither supercoiled or nicked open circle dsDNA substrates were degraded by gp60 (Figure 11C). These data demonstrate that Che9c gp60 has exonuclease activity similar to λ Exo and RecE.

Figure 11. *In vitro* assays demonstrate exonuclease activity of Che9c gp60.

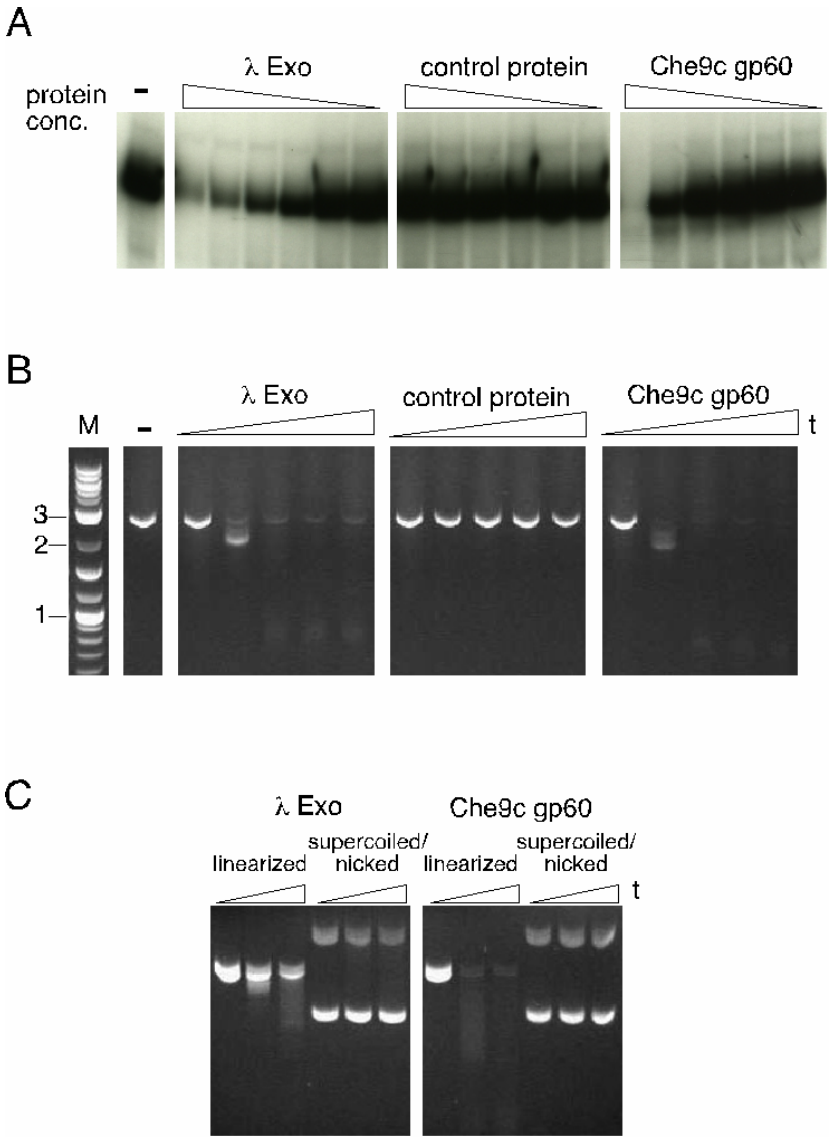


Figure 11. (A) Exonuclease activity was assayed by incubating Che9c gp60, λ Exo, or control protein extract with ^{32}P-labeled dsDNA (100 bp) for 5 minutes at room temperature, and the reactions analyzed by polyacrylamide gel electrophoresis. Reactions contained either no protein (–), or two-fold serial dilutions as indicated. Reactions with the highest protein concentrations contained Che9c gp60 at 0.2 μM or 5 U of λ Exo (NEB). The control protein extract was prepared from mock induced cells and the highest concentration corresponds to approximately 0.1 μg/ml. (B) Exonuclease activity of Che9c gp60 (0.2 μM), λ Exo (1 U/10 μl), or control protein was assayed in reactions with a 3 kbp linearized plasmid DNA substrate (0.8 nM), incubated for increasing amounts of time (t = 0, 2, 5, 7, or 10 minutes), and analyzed by agarose gel electrophoresis. The marker (M) indicates sizes in kbp. (C) Exonuclease activity of circular versus linear substrates was assayed. Che9c gp60 (final concentration 0.2 μM) or λ Exo (5 U) was incubated for increasing times (0, 5, and 10 min) similarly to (B) with a 3 kbp dsDNA substrate (2 nM) that was either supercoiled closed circular or linear (as indicated) and the products analyzed by agarose gel electrophoresis.

2.5 CHE9C GP61 BINDS ssDNA AND dsDNA

The λ Beta and *E. coli* RecT proteins both have numerous biochemical characteristics that distinguish them as SSAPs, including the formation of multimeric structures and the ability to bind both ssDNA and dsDNA and perform strand pairing, exchange and invasion. Several of these attributes were tested with the gp61 protein to determine if it acts similarly to RecT [227]. First, the DNA binding activities of gp61 were measured using a double-filter binding assay [20,239], and these results were confirmed by electrophoretic gel mobility shift assays. Similarly to RecT [68], gp61 binds ssDNA with moderate affinity (K_d = 163 ± 12.5 nM) and is only slightly reduced in binding affinity in the presence of Mg^{2+} (Figure 12A,B). Che9c gp61 bound dsDNA with a slightly lower affinity (K_d = 211 ± 4.2 nM), but this is substantially reduced with Mg^{2+} (Figure 12C,D), much like what is observed with RecT [68,69,145]. It is also of interest that gp61 bound ssDNA substrates at lengths of 20, 44, 48, and 76 nucleotides (nt) with similar affinities (Figure 12E). This is different than what is observed with λ Beta; gel shift assays have shown that Beta does not bind substrates that are 17 nt or 27 nt long, although it can bind to a 36mer [144]. Binding activity of gp61 was also observed by native polyacrylamide gel electrophoresis using both ssDNA and dsDNA substrates (Figure 12F), and quantification of the shifted bands reflects the binding affinities observed by filter binding analysis.

Figure 12. Che9c gp61 binds ssDNA and dsDNA.

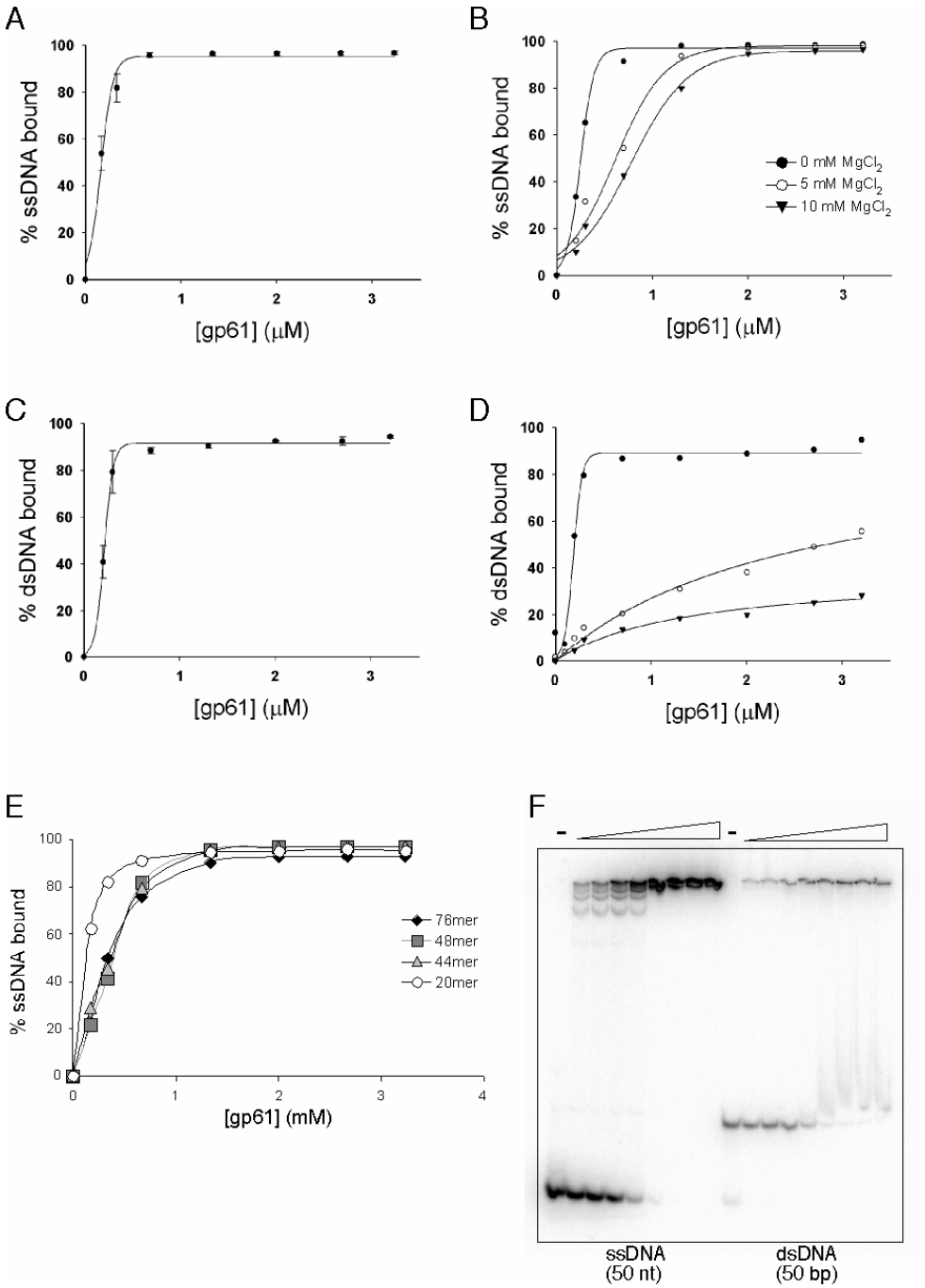


Figure 12. Purified gp61 protein at varying concentrations (0, 0.2, 0.3, 0.7, 1.3, 2.0, 2.7, 3.2 μM) was incubated with 66.7 nM ^{32}P-labeled ssDNA or dsDNA in binding assay buffer and analyzed either by double-filter binding assays (A-E) or native polyacrylamide gel electrophoresis (F). These experiments (without Mg^{2+}) were repeated in triplicate for both ssDNA (A) and dsDNA (C) and the data analyzed on SigmaPlot to determine binding constants. Reactions were also assayed with ssDNA (B) or dsDNA (D) containing 0 mM $MgCl_2$ (●), 5 mM $MgCl_2$ (○), or 10 mM $MgCl_2$ (▼). (E) ssDNA substrates of different lengths were tested (0 mM $MgCl_2$) and are depicted in the legend. (F) For gel shift assays, the same reactions from using either ssDNA (A) or dsDNA (C) were run on a native 8% polyacrylamide gel and analyzed.

It was observed that gp61-ssDNA complexes formed multiple distinct bands (at least four) in gel shift assays. These decreased in number as the concentration of protein was increased, and ultimately two large shifted bands were seen at a concentration of 2 μM gp61. This suggested that gp61 might form a multimeric complex upon binding to ssDNA, and is of importance since the formation of toroidal multimers is a property exhibited by other SSAPs such as λ Beta and RecT [162,224]. Large ring structures composed of up to 18 subunits are formed by λ Beta in the presence of ssDNA, and smaller rings (~12 subunits) are observed even the absence of DNA. Samples of gp61 were therefore prepared incubated with ssDNA substrates of several different lengths and analyzed these by electron microscopy. In the presence of even short ssDNAs (20 nt), gp61 formed small curved 'c-shaped' structures, although no structures were observed above background in the absence of DNA. As the length of the ssDNA increased, the size of the curved structures increased in diameter (Table 2), and many were circular in reactions with a 100 nt substrate (Figure 13A). The average diameter of the toroids formed by gp61 (14 – 16 nm) are similar to the diameter of the structures formed by λ Beta (18 - 21 nm) and RecT (18 nm) in the presence of ssDNA [162,224]. Both Beta and RecT also form helical filaments when bound to dsDNA [162,224], though this was not tested with gp61.

Table 2. Size analysis of Che9c gp61 structures observed by electron microscopy.

Length of ssDNA substrates (nt)	**20**	**44**	**48**	**76**
Average diameter of particles (nm)	9.74	12.17	13.48	15.74
Number of particles measured	10	14	13	11

Che9c gp61 protein was incubated with ssDNA substrates of varying lengths (20, 44, 48, 76 nt), stained with uranyl acetate, and visualized by transmission electron microscopy. Measurements were taken across the diameter of multiple particles and averaged.

Like λ Beta, the *E. coli* RecT protein also forms multimers in the absence of DNA that are visible by electron microscopy, although no structures could be seen with gp61 using simple negative staining. Gel filtration analysis of RecT (originally called p33) indicates that it forms a tetramer [68]. Therefore, analytical gel filtration was used to determine the state of gp61 in solution. Three concentrations of recombinant gp61 protein were run on a Superdex gel filtration column that had been standardized with both low and high molecular weight proteins. As the concentration of gp61 was increased from 5 μM to 25 μM, the size of the complex increased. At 5 μM, gp61 eluted at a time corresponding to approximately 70 kDa, which is roughly twice the size of the predicted molecular weight of gp61 (40 kDa). At 10 μM, it eluted at 102 kDa, and at 25 μM it eluted at 143 kDa. Although these data do not fit exactly with the predicted size of multimers of gp61 (e.g., 80 kDa, 120 kDa, or 160 kDa), the native molecular weight of the standards varied slightly on this column (±14 kDa) (Figure 13B). Thus, it appears that at the highest concentration tested, gp61 likely forms a tetramer, much like what has been observed for RecT (concentration not given for RecT experiment; [68]). Additionally, increasing the salt (NaCl) concentration in the buffer from 100 mM to 300 mM did not change the size of the gp61 complexes eluted by gel filtration (data not shown). This indicates that the multimerization of gp61 in solution is not likely the result of non-specific protein-protein interactions; further anlaysis would be required to completely rule out this possibility. Reactions containing gp61 incubated with ssDNA eluted in the void volume, indicating that they were larger than the pore size of this column, which is consistent with gp61-ssDNA complex formation (data not shown). Finally, while these data are not conclusive, they support the hypothesis that gp61 forms multimers in the absence of DNA.

Figure 13. Multimeric structures formed by gp61 in the presence and absence of DNA.

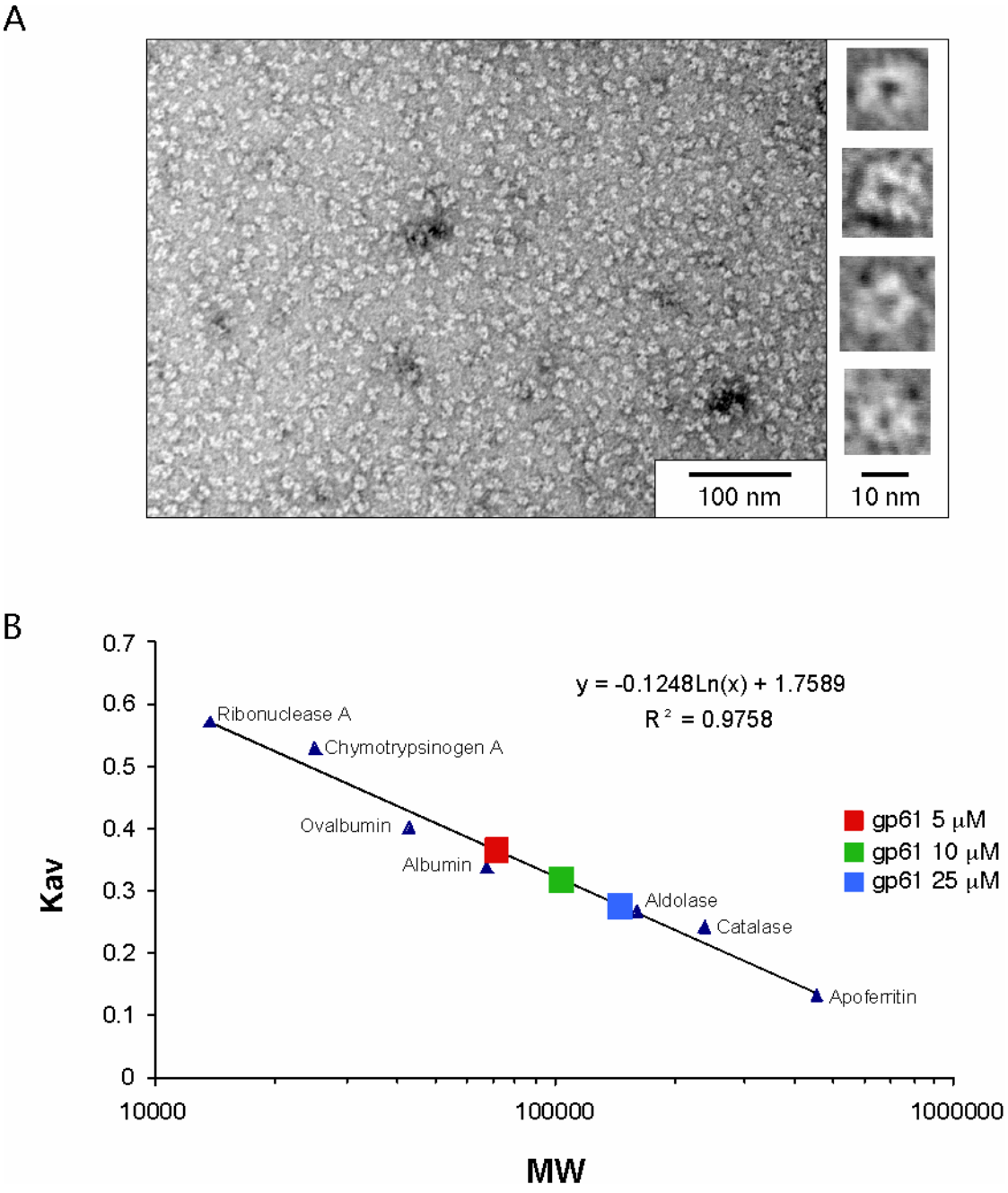


Figure 13. (A) Electron micrograph depicting Che9c gp61 protein multimers in the presence of ssDNA. Reactions containing gp61 protein (1.2 μM) incubated with ssDNA (100 nt; 1.9 μM) were absorbed to copper grids, stained with 2% uranyl acetate and examined by transmission electron microscopy. Images were collected at a magnification of 140,000x; four examples of toroid structures are shown alongside a size bar for reference. (B) Protein standards (high and low molecular weight) were run on a Sephadex high-performance gel filtration column, elution times recorded, and the K_{av} value was determined for each standard. These were plotted against the molecular weight (on a logarithmic scale) and a trendline wasfit to the data; the equation and fit value are depicted on the graph. Using this equation, the elution times for each gp61 sample (5, 10, and 25 μM) were calculated to determine the molecular weight of the native protein complex, and these were graphed on the trendline.

2.6 CONCLUSIONS

The mosaic architecture commonly observed in bacteriophages is exemplified by phage-encoded SSAPs and their cognate exonucleases [85] in that there is no apparent consistency with the specific pairing of these proteins. Further, SSAP-exonuclease recombination systems are rare in sequenced mycobacteriophages. Bioinformatic analyses identified mycobacteriophage Che9c gp60 and gp61 as homologues of the *E. coli* Rac prophage RecE and RecT proteins. Finding these genes in mycobacteriophages was somewhat suprising; previous analyses found RecT-like proteins predominantly in the low GC% Gram-positive bacteria [85], while the mycobacteria have a high G+C% (~65-67%) [22]. However, the Che9c-encoded proteins were distantly related to the *E. coli* RecET proteins with less than 30% amino acid identity conserved. Similarly to the organization of these genes seen in other phages, the mycobacteriophage-encoded SSAP is located next to a predicted exonuclease.

Desite the weak identity observed for these proteins, the biochemical properties of Che9c gp60/gp61 support the finding that they are RecET homologues. Che9c gp60 has dsDNA exonuclease activity that does not act on circular substrates, which has been observed for RecE. Che9c gp61 binds ssDNA and dsDNA with affinities and Mg^{2+} dependencies similar to RecT, and it forms toroids in the presence of ssDNA. Interestingly, gp61 binds to ssDNA substrates as small as 20 nt with higher affinity than longer substrates. Although it is reported that Beta cannot bind substrates shorter than 36 nucleotides [144], more recent experiments indicate that it does indeed bind smaller lengths of ssDNA (D. Court, personal communication).

Numerous additional *in vitro* assays could be performed with each of these proteins to further characterize their activities, such as strand pairing and strand exchange assays for gp61. These assays were attempted but were not successful for control reactions. It would also be

helpful to analyze gp61 by native gradient gel electrohoresis to confirm the gel filtration data which suggest it is a trimer or tetramer in its native state. Also, RecT DNA binding is affected by salt (NaCl) concentrations greater than 50 μM [145]. It was suggested that the varying salt and Mg^{2+} requirements for ssDNA and dsDNA binding may reflect different types of binding [145], which is supported by the observation that ssDNA- and dsDNA-gp61 complexes form toroids and helical filaments, respectively [224]. Additional experiments could therefore be performed with gp61 in buffers containing Mg^{2+} and/or NaCl to determine their effect on multimer formation. $MgCl_2$ is absolutely required for formation of λ Beta multimers in the absence of DNA, and addition of Mg^{2+} to reactions with Beta and ssDNA appears to stabilize the formation of large rings [162]. However, Beta also requires Mg^{2+} for ssDNA binding [144], whereas RecT and gp61 do not [145]. In fact, RecT multimers are inhibited with high concentrations of $MgCl_2$, though a low concentration (0.3 mM) is required for the formation of small circles but not rod-like structures [224]. Therefore varying Mg^{2+} concentration or testing other cations might enhance gp61 multimerization, and this could be assayed in the future.

Collectively, the biochemical data clearly demonstrate that gp60 and gp61 function equivalently to the *E. coli* λ Red and RecET systems in all assays tested. Thus, these proteins have properties consistent with their utility as a means to develop a recombineering system in the mycobacteria using mycobacteriophage-encoded proteins.

3.0 DEVELOPMENT OF THE MYCOBACTERIAL RECOMBINEERING SYSTEM

3.1 INTRODUCTION

Recombineering is a widely used system for mutagenesis in *E. coli* [38], which has recently been extended to other Gram-negative bacteria, such as *Salmonella* and *Shigella* [42,185]. A variety of techniques can be performed with ease including gene replacements, point mutations, deletions, small insertions, *in vivo* cloning, and modifications of bacterial artificial chromosomes and genomic libraries [36,38,199]. Recombineering exploits the potent recombination activities of the λ Red proteins and is a simple means of increasing homologous recombination in the host bacterium with minimal DNA manipulations or screening.

Recombineering is most commonly used for construction of allelic gene replacements and point mutations on either the bacterial chromosome or on BACs. Substrates for gene replacements contain an antiobiotic resistance gene of choice (and a *sacB* gene for counter-selection if desired) with short lengths of homology to the gene target on either end (Figure 7) [223]. These strains can be easily unmarked with ssDNA substrates that have homology flanking the antibiotic and *sacB* cassettes with selection for sucrose resistance. Optimum numbers of recombinants (up to 10^5 out of 10^8 viable cells) are obtained with 100 – 300 ng of the AES. Recombination of the dsDNA AES substrates is dependent on Exo and Beta, and the addition of Gam increases recombineering efficiencies significantly [135,240].

Mutagenesis with ssDNA substrates is a particularly useful strategy that permits numerous types of mutations to be introduced into bacterial and phage genomes [52,151]. This recombination only requires the Beta protein, although some experiments have shown that the presence of Exo (even with ssDNA substrates) can slightly enhance recombination [244]. Maximal recombination frequencies are achieved with 70 nt ssDNAs [52], and these substrates are designed with the mutation(s) centered such that they anneal to the lagging strand for DNA replication, since these are more efficiently recombined [52].

The *E. coli* recombineering strains that are typically used express the λ Red/Gam proteins under control of the temperature-sensitive repressor CI857 in a defective λ prophage. Following mutagenesis, the prophage can be removed or the mutation can be moved to a clean background by generalized transduction with phage P1 [223]. Alternatively, there are various versions of recombineering plasmids with temperature-senstive replicons that can be transferred into different strains and quickly removed [42]. These appear to work just as well as the prophage version, and can be used in other Gram-negative bacteria. Recombineering strains are induced at mid-logarithmic phase for 15 minutes and prepared for transformation; this expression time is optimal for recombineering but is short enough that cell death does not occur from the toxic effects of Gam [240].

Clearly, the strategic use of the λ Red proteins to develop the widely-applicable recombineering technology in *E. coli* sets a precedent for improving genetic techniques in other bacterial systems. Mycobacterial researchers in particular would benefit enormously from a simple and efficient system for genetics – such as recombineering – that alleviates time-consuming DNA manipulations and screening procedures. However, the λ Red proteins, while useful for *E. coli* and related Gram-negative organisms, reportedly are not functional in the

mycobacteria, which is not surprising considering the divergence between these two groups of bacteria. From the bioinformatic and biochemical data, the mycobacteriophage-encoded RecET homologues appeared to be the most likely candidates for developing a recombineering system for the mycobacteria. Accordingly, expression of the Che9c proteins gp60 and gp61 elevates homologous recombination in both fast- and slow-growing mycobacteria. This facilitates the use of recombineering-based strategies for mutagenesis of mycobacterial and mycobacteriophage genomes. This chapter will discuss the development of the Che9c-based mycobacterial recombineering system and its various applications.

3.2 EXPRESSION OF CHE9C RECOMBINATION GENES *IN VIVO*

In order to test the activity of Che9c gp60 and gp61 *in vivo*, various plasmids were constructed that express these genes in the mycobacteria (see Table 15) [227]. Although genes *59* and *62* encode small proteins, it was possible that one of these might encode a Gam-like functional protein that is not identifiable by bioinformatic analysis. Therefore, this region of Che9c was cloned under the control of the *M. smegmatis* acetamidase promoter ($P_{acetamidase}$) [44,157] on an extrachromosomally-replicating parent vector (pLAM12) to create pJV24 (Figure 14). The acetamidase promoter is comprised of the upstream region of the *M. smegmatis* acetamidase gene regulon and terminates at the start codon of the acetamidase gene (*amiE*) [157,190]. Therefore, placing the start codon of a gene at this locus results in translation from the ribosome binding site (RBS) encoded by this cassette. Additional constructs were made in the parent vector pLAM12 containing only genes *60* and *61*; in this case, gene *60* was under translational control from either its endogenous signals (pJV53; Figure 14) or those present in the acetamidase

promoter cassette (pJV63). Similar plasmids were constructed containing only gene *60* (pJV61 and pJV55) or gene *61* (pJV52 and pJV62), either under their own translation signals or those of the acetamidase cassette, respectively. All of these plasmids were transformed into *M. smegmatis* for further assays. Genes *59-62* were also cloned under control of the constitutive *M. bovis* BCG *hsp60* promoter (pJV23), but this plasmid did not produce transformants in *M. smegmatis.* It is also noted that *M. smegmatis:pJV63* grows slowly and does not grow on plates in the presence of inducer (data not shown). This may be due to the somewhat toxic effects of gp60, since strains expressing only this protein (even in the absence of induction) grow slowly compared to empty vector control strains (data not shown).

Figure 14. Mycobacterial plasmids expressing Che9c genes.

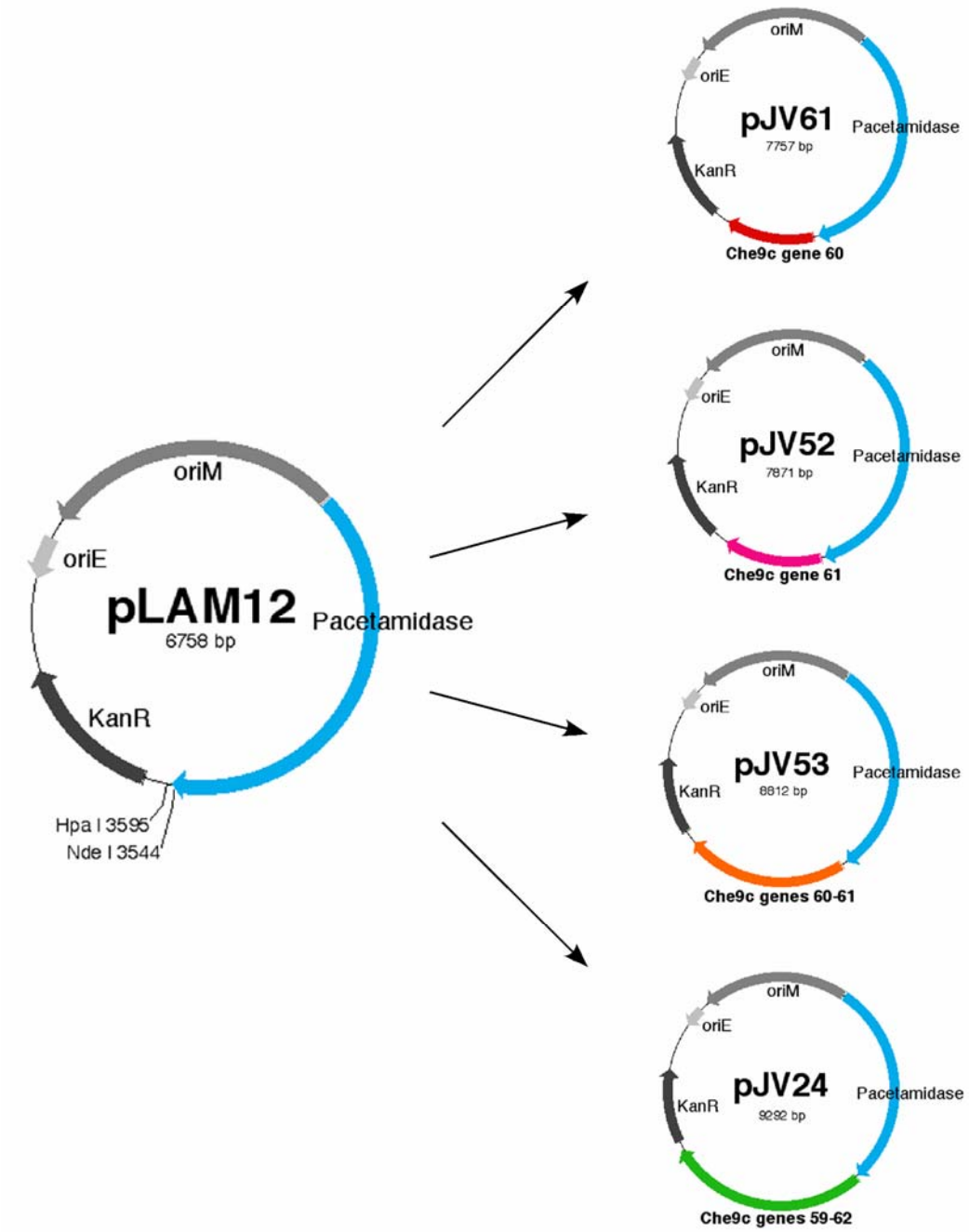


Figure 14. Plasmid pLAM12 is an extrachromosomally-replicating plasmid that contains a kanamycin-resistance gene and the acetamidase expression cassette ($P_{acetamidase}$), which has an inducible promoter and translation signals (ribosome binding site: RBS); placing the start codon of a gene at site *Nde*I results in a translational fusion to this RBS. For the plasmids shown, Che9c genes *60* and *61* were cloned separately or together downstream of $P_{acetamidase}$ into the *Hpa*I site with their endogenous RBSs included. Plasmid pJV24 was constructed similarly but includes genes *59-62*. Several plasmids were constructed similarly to those shown by placing the Che9c genes at the *Nde*I site for translational fusion; these are not depicted.

Protein expression was monitored by western blot analysis for several of these strains with polyclonal antibodies generated against purified gp61 protein (Figure 15) [227]. All strains in which gene *61* was under control of $P_{acetamidase}$ had detectable expression of gp61 after three hr of induction with acetamide. It was seen that some strains expressed more gp61 than others, although there was no correlation between expression levels and whether the endogenous translation signals or signals from the acetamidase cassette were used (Figure 15 A,B; compare strains with pJV53 to pJV63, and pJV52 to pJV62, respectively). Specifically, a strain with pJV62 ($P_{acetamidase}$ RBS) had three-fold less gp61 expression than pJV52 (endogenous RBS), whereas the opposite was true for pJV53 (endogenous RBS) and pJV63 ($P_{acetamidase}$ RBS). The strain expressing gp59-gp62 (*M. smegmatis:pJV24*) consistently showed expression of gp61 in the absence of acetamide (Figure 15A,C), and this may be due to leaky expression sometimes observed with this promoter in succinate medium [157]. This level of protein expression was not observed with any other *M. smegmatis* strain, and further there is no expression observed in media containing ADC (see Materials and Methods) that is reported to be repressive of this promoter (Figure 15D). Strains of *M. tuberculosis* containing the same plasmids were also tested for protein expression, and it was observed that there was much more leaky expression of the promoter in the absence of induction (Figure 15E).

Figure 15. Western blot analysis of mycobacterial strains expressing Che9c proteins.

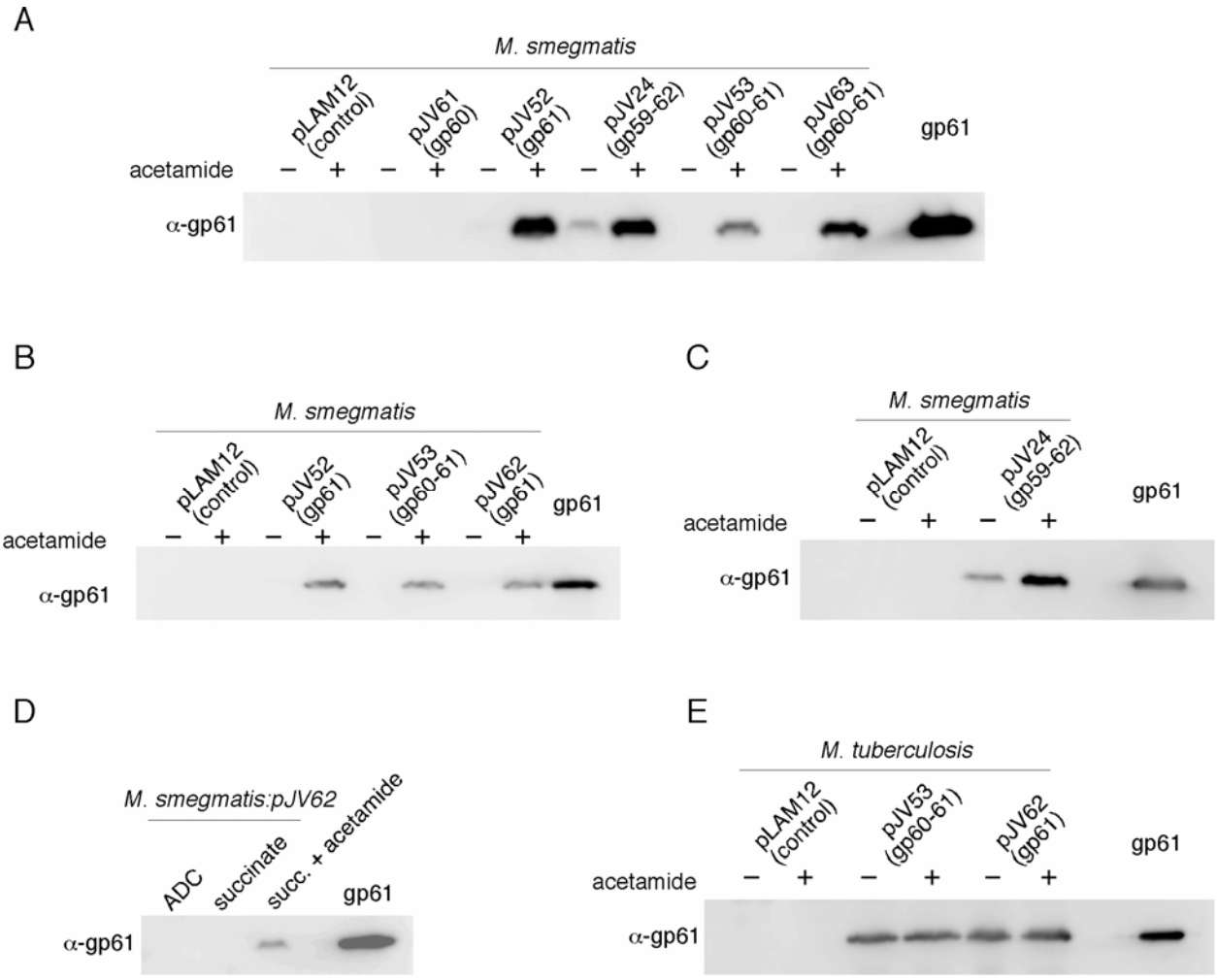


Figure 15. Strains of either *M. smegmatis* (A-D) or *M. tuberculosis* (E) containing various plasmids were grown to mid-log phase and samples split; one culture was induced with 0.2% acetamide, and both were grown for 3 hours. Cell aliquots were normalized to OD_{600} and samples were run on SDS polyacylamide gels and analyzed by western blot with polyclonal anti-gp61 antibodies.

Cultures of *M. smegmatis* mc^{2}155:pJV24 repeatedly showed a slight decrease in viability (assayed by colony counts) after four hours of induction with acetamide that continued to decline up to 24 hours, whereas mc^{2}155:pJV53 did not (Figure 16A and data not shown). This may be due to the increased levels of protein expression in mc^{2}155:pJV24, as seen by western blot (Figure 16B), or alternatively could result from the expression of Che9c gp59 and/or gp62. The strain containing the empty control vector (mc^{2}155:pLAM12) surprisingly grows more slowly than strains expressing Che9c genes. Ultimately, three hours of induction appeared to give adequate levels of protein expression without any potential toxic effects, and this also is approximately the length of time required for *M. smegmatis* doubling in this media. Therefore, these strains were tested for recombination activity *in vivo* using this induction procedure.

Figure 16. Growth curves and expression profiles of strains expressing Che9c gp60 and gp61.

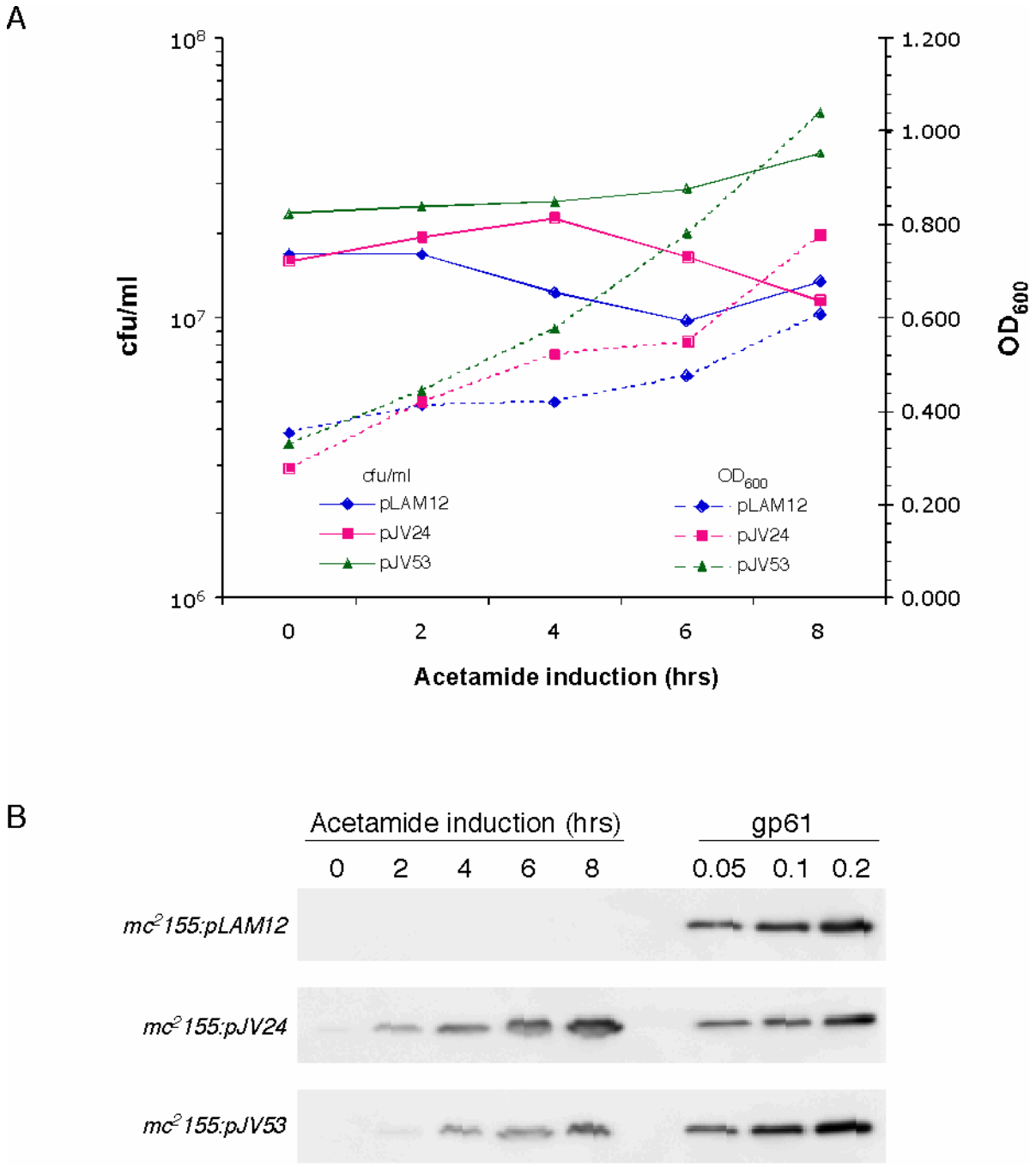


Figure 16. *M. smegmatis* strains containing plasmids pLAM12 (empty vector control), pJV24 (Che9c gp59-62), and pJV53 (Che9c gp60-61) were grown to mid-log phase and induced with 0.2% acetmide (time point 0 hours). (A) Cells were plated to determine viability (cfu/ml) and absorbance (OD_{600}) readings taken every two hours. (B) Aliquots of each culture were removed at each time point, normalized to OD_{600}, and analyzed by western blot analysis with antibodies against gp61.

3.3 ALLELIC REPLACEMENT MUTAGENESIS

3.3.1 Che9c gp60 and gp61 promote homologous recombination *in vivo*

To determine if Che9c gp60 and gp61 can function *in vivo* to promote elevated levels of homologous recombination, *M. smegmatis* strains expressing these genes were transformed with a linearized AES targeting the *leuD* gene. Deletion of this gene confers leucine auxotrophy [14,81] and therefore facilitates a phenotypic assay for correctly targeted genes; growth medium without leucine only supports growth of recombinant colonies that are not correctly targeted for gene replacement. The AES tested contained ~1000 bp of homology to the *leuD* locus flanking hyg^R and *sacB* genes (Figure 17A). Strains *M. smegmatis* mc^2155:pJV24 (expressing Che9c gp59-62) and *M. smegmatis* mc^2155:pLAM12 (empty vector control) were transformed with 100 ng of the *leuD* AES, and the reaction was split onto media with or without leucine. Hyg^R colonies (43) were recovered on media containing leucine and only in the strain strain expressing Che9c gp60 and gp61, while no colonies were obtained on media lacking leucine or in the control strain (Figure 17B). This indicates that expression of Che9c genes increases homologous recombination above background levels and that each recombinant colony obtained was the result of a correctly targeted allelic exchange [227].

Figure 17. Allelic gene replacement of the *M. smegmatis leuD* gene.

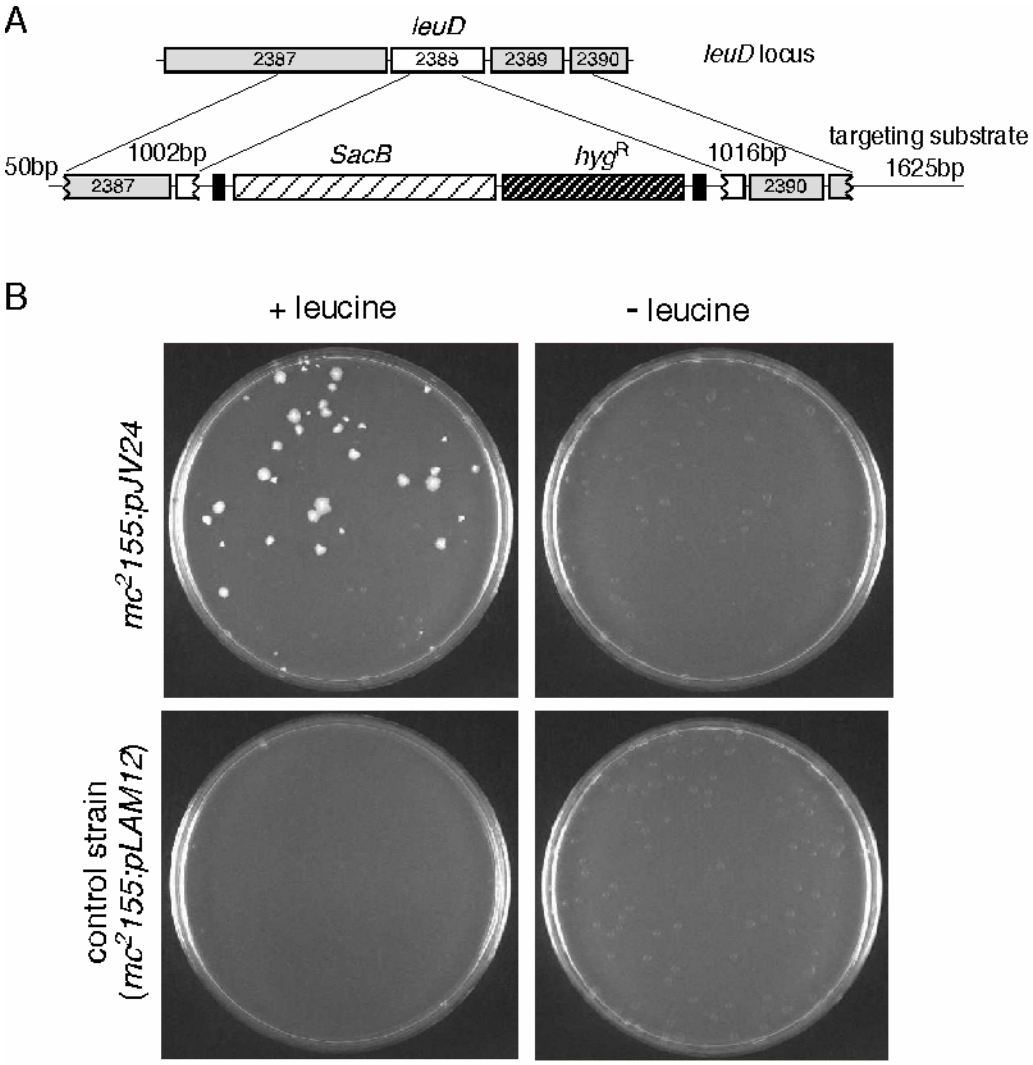


Figure 17. [227] (A) An AES targeting *M. smegmatis leuD* is depicted in this schematic; plasmid p0004S:leuD contains ~1000 bp of homology flanking a *hyg*^{R} and *sacB* gene, and this was linearized by restriction digest. (B) Strains *M. smegmatis* mc^{2}155:pJV24 and *M. smegmatis* mc^{2}155:pLAM12 were grown to mid-logarithmic phase, induced with acetamide for three hours, and electrocompetent cells prepared. 100 ng of the *leuD* AES were transformed into these strains, recovered for four hours, and the reaction split onto media in the presence or absence of leucine.

3.3.2 Recombineering requires both Che9c gp60 and gp61.

A similar assay targeting the *M. smegmatis leuB* gene was used to dissect the genetic requirements for recombination, and this demonstrated that expression of both Che9c gp60 and gp61 is necessary and sufficient for recombineering (Table 3) [227]. The presence of genes *59* and *62* in plasmid pJV24 repeatedly yielded similar recombineering frequencies to plasmid pJV53, which expresses only genes *60* and *61* (Table 3 and Figure 19), and these two strains gave the highest recombineering frequencies. Interestingly, although pJV63 produces higher levels of protein expression than pJV53 (Figure 15A), this strain was reduced for recombination activity. Based on these data, strain mc^2155:pJV53 was used for most subsequent experiments because it does not exhibit a viability defect phenotype.

Table 3. Recombineering requires both Che9c gp60 and gp61.

Strain (proteins encoded)[a]	Recovered colonies w/leucine [b]	Recovered colonies w/o leucine[c]	Cell competency[d] (cfu/μg DNA)	Recombineering frequency[e] (w/leucine)
mc^2155:pLAM12 (control strain)	0	1	5.8×10^5	0
mc^2155:pJV61 (gp60 only)	0	0	1.2×10^6	0
mc^2155:pJV52 (gp61 only)	0	0	6.0×10^5	0
mc^2155:pJV24 (gp59-62)	52	0	6.4×10^5	1.6×10^{-3}
mc^2155:pJV53 (gp60-61)	57	1	1.4×10^6	8.3×10^{-4}
mc^2155:pJV63 (gp60-61)[f]	7	0	4.8×10^5	2.9×10^{-4}

a. Each strain contains an extrachromosomally-replicating plasmid expressing varying combinations of Che9c gp60 and gp61.
b,c. Cells were transformed with 100 ng of an AES targeting *leuB*, and recovered cells were split on media with or without leucine supplement.
d. Cell competency is determined as the cfu/μg plasmid pPGA1, an integration-proficient vector providing hygromycin resistance, when 50 ng was transformed.
e. Recombineering frequency (recombinant cfu/μg DNA/cell competency) is shown for transformations with the *leuB* substrate (p0004S:leuB) and that are plated on leucine-supplemented media.
f. mc^2155:pJV63 expresses Che9c gp60/gp61 under control of the acetamidase promoter through a translational fusion to that cassette, in contrast to mc^2155:pJV53 in which these genes are expressed from their endogenous signals.

3.3.3 Recombineering of the *M. smegmatis groEL1* gene

Recombineering of other loci had similar results to those obtained at the *leuD* and *leuB* loci, and genotypic analysis of the recombinants demonstrated that 90% or greater were correctly targeted [227]. First, the *groEL1* gene was targeted using an AES with ~500 bp of homology on each end (Figure 18A), which was amplified by PCR using a circular AES as a template (see Figure 21A). Colony PCR analysis shows that each of the ten colonies tested in this example were allelic gene replacements of *groEL1* (Figure 18B), and Southern blot analysis confirmed these results (Figure 18C). Additionally, several Δ*groEL1* mutant strains constructed by recombineering exhibited the expected biofilm defects for this strain (data not shown) [149]. Shortening the homology lengths of the *groEL1* AES resulted in a decrease in recombination, such that less than ten colonies were obtained with 50 bp homology regions (Figure 19), and only ~ 50% of these were correctly targeted (data not shown). Not surprisingly, there is a low level of recombination activity in the absence of induction due to leaky expression from the acetamidase promoter, which has been observed in these experiments and others (data not shown, and K. Derbyshire, personal communication). Extending the induction time from three hours up to ten hours only slightly increased recombineering frequencies (less than two-fold, data not shown); therefore a three-hour induction was used for all subsequent experiments. This recombination activity is somewhat dependent on host RecA, since recombination frequencies were decreased five-fold in an *M. smegmatis* Δ*recA* strain (Table 4); the effect of *M. smegmatis* RecA dependence was small compared to the 10-50 fold decrease observed in *E. coli* in recombineering assays [135,240].

Figure 18. Allelic gene replacement of the *M. smegmatis groEL1* gene.

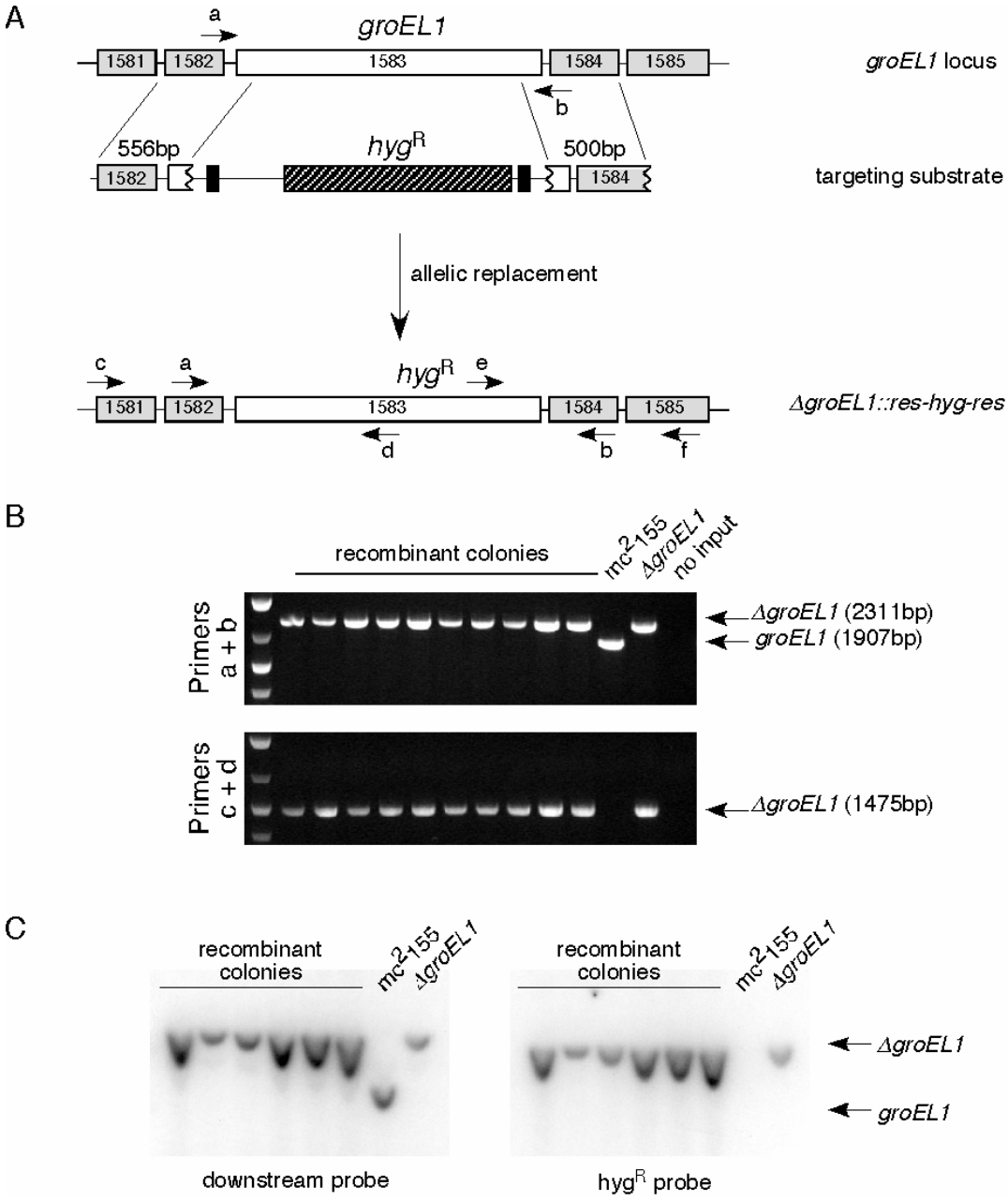


Figure 18. [227] (A) The *groEL1* AES was generated by cloning approximately 500 bp of homology flanking a *hyg*^{R} gene (plasmid pMsgroEL1KO; see Figure 21) and PCR amplifying the region shown. Homologous recombination of this AES with the *groEL1* locus results in allelic exchange as shown. The locations of primers a, b, c, and d are shown (e and f are depicted for assays shown in Figure 21) [JCV67+68, JCV71+94, JCV72+172]. (B) Colony PCRs from recombinant colonies using primer pairs a and b (1.9 kbp wild type, 2.3 kbp mutant *groEL1:res-hyg-res*), and c and d (no product for wild type, 1.5 kbp mutant *groEL1:res-hyg-res*) are shown; c and d are present in the chromosome of recombinants only. DNA from wild type *M. smegmatis* or Δ*groEL1* mutant strains were used as controls. (C) Southern blot analysis of DNA isolated from gene replacement mutants using either a probe to the downstream homologous region of the *groEL1* locus, or a probe to the *hyg*^{R} gene. Expected band sizes: 2.3 kbp wild type, 3.3 kbp mutant *groEL1:res-hyg-res*; DNA from wild type *M. smegmatis* or Δ*groEL1* mutant strains were used as controls.

Figure 19. dsDNA recombineering dependence on homology length.

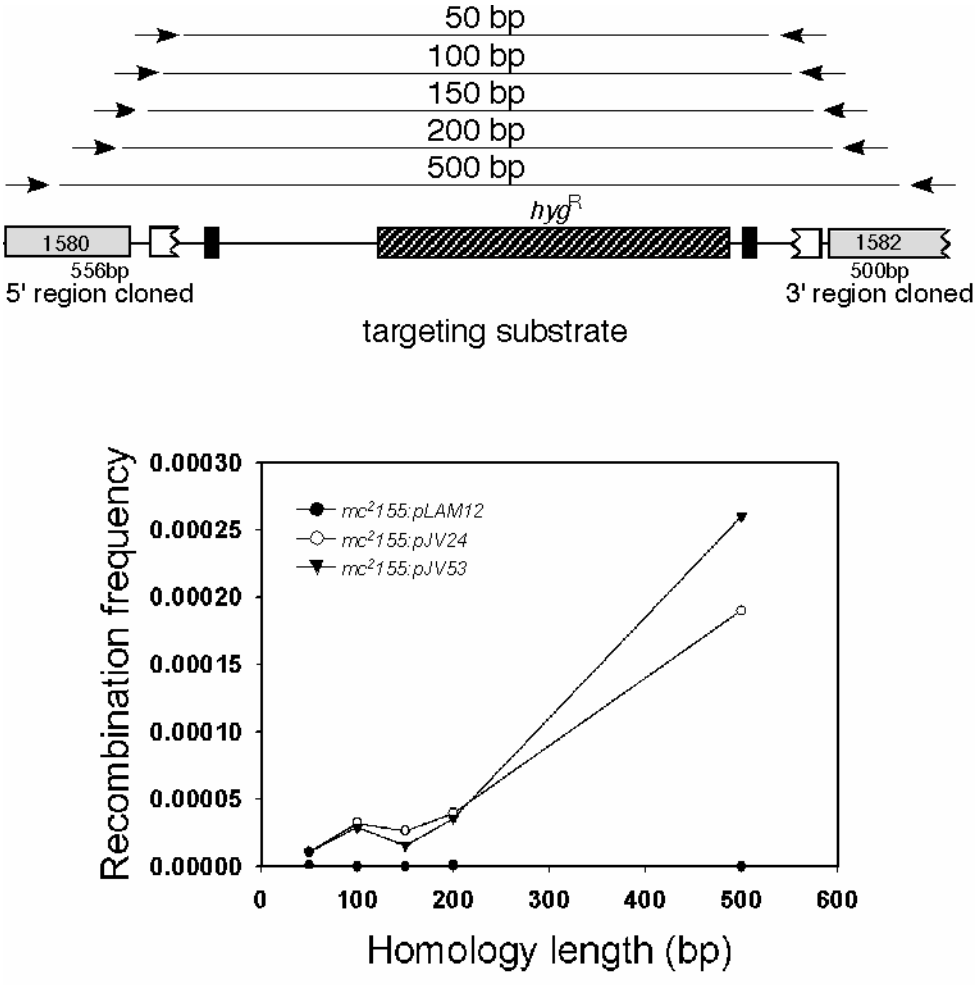


Figure 19 [227]. Plasmid pMsgroEL1KO contains 556 bp and 500 bp of homology 5′ and 3′ of the *groEL1* gene, respectively, flanking a *hyg*^{R} gene. Primer pairs were designed to amplify this region resulting in PCR products with homology lengths of 50 bp, 100 bp, 150 bp, 200 bp, and 500 bp. These substrates were transformed into *M. smegmatis* strains containing plasmids pLAM12 (●), pJV24 (○), and pJV53 (▼), and recombineering frequencies are shown on the y-axis.

Table 4. dsDNA recombineering dependence on host RecA.

Strain	Recovered colonies with *groEL1* AES[b]	Cell competency (cfu/μg)[c]	Recombineering frequency[d]
mc^2155:pJV53	226	6.0×10^6	3.8×10^{-4}
mc^2155:pJV53 Δ*recA*[a]	99	1.3×10^7	7.6×10^{-5}

a. The *M. smegmatis ΔrecA* strain was constructed by allelic gene replacement by recombineering and unmarked using γδ resolvase, as described in the Materials and Methods.
b. Electrocompetent cells of the two strains were transformed with 100 ng of the *groEL1* AES (see Figure 18), and Hyg^R colonies were recovered; the data represent the average of two experiments.
c. Cell competency is determined as the cfu/μg plasmid pJV39, an integration-proficient vector providing hygromycin resistance, when 50 ng was transformed.
d. Recombineering frequency is calculated as the number of recombinant cfu per μg DNA divided by the cell competency.

3.3.4 Recombineering frequencies are limited by DNA uptake efficiency.

Using 100 ng of the dsDNA substrates for allelic exchange typically produced between 50 and 200 recombinant colonies, and the number of colonies obtained by recombineering was directly proportional to the ability of the electrocompetent cells to productively take up DNA (referred to as 'cell competency'). Control transformations with an integration-proficient plasmid were performed with 50 ng to determine cell competency, and these values are reported as transformants (colony forming units; cfu) per μg DNA (Table 3, Table 4, and Table 5), which is typically ~10^6 cfu/μg. Therefore a 'recombineering frequency' is used to compare experiments; this is calculated as the number of recombineering transformants per μg DNA divided by the cell competency. When using 100 ng of the AES, recombineering frequencies averaged 1-5 x 10^{-4} (Table 5). Increasing the amount of AES (up to one μg) does not result in a higher recombineering frequency (data not shown), but rather this can be accomplished by increasing the competency of the cells by optimizing the protocol for electrocompetent cell preparation (see Materials and Methods).

3.3.5 Recombineering of other *M. smegmatis* genes

Several additional *M. smegmatis* loci were also tested for targeted gene replacement and yielded similar recombineering frequencies to *groEL1* [227]. The number of colonies recovered for each gene locus were comparable, and again more than 90% were correctly targeted (Table 5). Frequencies were observed to vary between 10^{-5} and 10^{-4}, with even higher frequencies (10^{-3}) from targeting the *leuD* and *leuB* genes (Table 3 and Table 5). This is likely a result of the longer homology lengths utilized in these experiments and this corroborates the observation that increasing homology length increases recombineering frequencies (Figure 19). Occasionally, recombination frequencies at these two loci were more similar to those for other loci such as *groEL1* (compare *leuB* in Tables 3 and 5).

Table 5. Recombineering of *M. smegmatis* loci.

Gene targeted[a,b]	Recovered colonies[c]	Recombination frequency[d]	Gene replacements[e]
0651	478	4.8×10^{-5}	>90%
1583 (*groEL1*)	180	1.8×10^{-4}	>90%
2379 (*leuB*)	25	1.9×10^{-4}	>90%
2388 (*leuD*)	43	1.2×10^{-3}	>90%
2723 (*recA*)	128	1.3×10^{-4}	90%
4303	281	2.8×10^{-5}	>90%
6048 (*cobW*)	280	2.8×10^{-4}	90%
6065 – 6067	242	2.4×10^{-5}	ND
6067 - 6068	280	2.8×10^{-4}	ND

a. Genes were targeted using linearized plasmid DNA substrates (digested with two enzymes adjacent to the homologous sequences and *oriE* region removed; see text below) containing a Hyg^R cassette flanked by ~500 bp homology to the locus.
b. The gene locus number is the new locus tag (MSMEG_XXXX).
c. *M. smegmatis* mc^2155:pJV24 cells were transformed with 100 ng of each targeting substrate and Hyg^R colonies recovered; cfu for *leuB* and *leuD* represent half of the transformation plated on leucine supplemented media.
d. Recombination frequencies are represented as recombinant cfu per μg divided by cell competency, in which the transformation efficiency is determined by using a Hyg^R, integration-proficient vector. For all loci except *leuD* and *leuB*, the cell competency was 1×10^7 cfu/μg; for *leuB*: 1.3×10^6 cfu/μg; for *leuD*: 7.2×10^5 cfu/μg.
e. The number of correctly targeted gene replacements was determined by PCR or phenotypic analysis (*leuB* and *leuD*), with a minimum of 10 colonies each.

3.3.6 Recombineering of the *M. tuberculosis groEL1* gene

Since recombineering was successful in *M. smegmatis*, the effectiveness of this system was tested in *M. tuberculosis* by targeting the *groEL1* gene (Figure 20A) [227]. The results were similar to those seen in *M. smegmatis*; ~150 recombinant colonies were obtained in an *M. tuberculosis* H37Rv:pJV53 strain (Figure 20B and Table 6), yielding a recombineering frequency of 1.7×10^{-4}. Out of 16 colonies examined by Southern blot analysis, at least 14 were correctly targeted to the *groEL1* locus (Figure 20C). Although colonies were observed in a control strain (H37Rv:pLAM12), these grew slowly (Figure 20B), arose at a much lower frequency (8.1×10^{-6}), and none were correctly targeted when examined by Southern blot analysis (Figure 20D). However, the *M. tuberculosis* H37Rv:pJV53 strain showed protein expression in the absence of acetamide induction (Figure 15E), which is not observed in *M. smegmatis* (Figure 15A). This is not surprising, since previous studies have shown that *M. tuberculosis* has a lower tolerance than *M. smegmatis* for plasmids containing the acetamidase promoter cassette [25]. Therefore, as an additional experiment, cultures were grown in OADC (see Materials and Methods), washed, and grown for 24 hours in media containing succinate and acetamide prior to harvesting for electrocompetent cells. Recombinants were also correctly targeted in these experiments (Figure 20C,D), and recombineering frequencies were similar; however, cell competency was overall lower for these strains. Although protocols for preparing electrocompetent cells of *M. tuberculosis* do not recommend storing cell aliquots at -80°C, this did not have an effect on the overall recombineering frequency. However, freezing cells did lower cell competency (Table 6), which has been observed previously [80]. Additionally, using a

PCR-generated *groEL1* AES yielded approximately five-fold more recombinants than a pMtbgroEL1 plasmid AES linearized by restriction digest (Table 6), and a higher proportion of the PCR colonies were correctly targeted (Figure 20D). The incorrect targeting of the *Pac*I-digested *groEL1* AES was likely due to targeting to the pJV53 plasmid via the homology at the *oriE* region (see below). Using PCR-generated substrates also reduced the background in the control strain (Table 6).

Table 6. Recombineering frequencies from targeted gene replacement of the *M. tuberculosis groEL1*.

	Recombineering frequencies with each AES[b]		
Strain and growth media[a]	pYUB854 *Pac*I-digest	pMtbgroEL1KO *Pac*I-digest	pMtbgroEL1KO PCR
H37Rv:pLAM12, OADC	1.7×10^{-5}	3.3×10^{-5}	5.6×10^{-6}
H37Rv:pLAM12, succinate	1.3×10^{-5}	1.2×10^{-5}	8.1×10^{-6}
H37Rv:pJV53, OADC	3.3×10^{-5}	3.3×10^{-5}	2.1×10^{-4}
H37Rv:pJV53, succinate	2.7×10^{-5}	3.7×10^{-5}	1.7×10^{-4}
H37Rv:pJV53, frozen aliquots[c]	ND	ND	4.3×10^{-4}

a. Strains were grown to mid-logarithmic phase in media and induced for 24 hours as follows: either grown initially to mid-logarithmic phase in media (1) containing OADC, washed, and induced in succinate and acetamide media, or (2) containing succinate, and subsequently acetamide added for induction.
b. The AESs were prepared either by restriction digest with *Pac*I of plasmids pYUB854 (no homology) or pMtbgroEL1KO (*groEL1*), or by PCR-amplification.
c. Electrocompetent cell aliquots (cells grown in succinate media) were frozen at -80°C for two weeks, thawed, and transformed with the PCR-amplified *groEL1* AES.

Figure 20. Allelic replacement of the *M. tuberculosis groEL1* gene by recombineering.

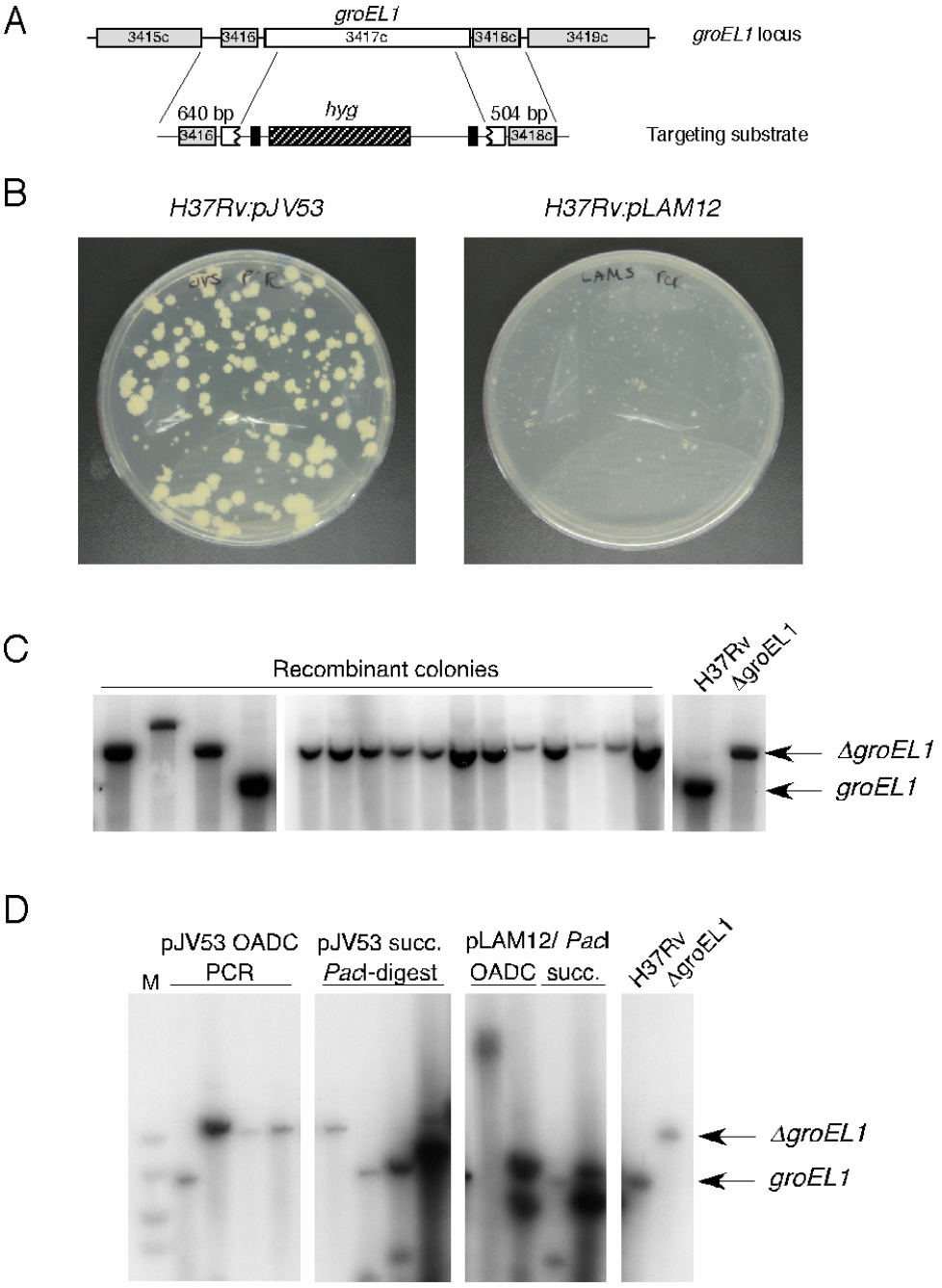


Figure 20 [227]. (A) Schmetic of the *M. tuberculosis groEL1* gene locus and the *groEL1* AES. (B) Pictures of the recovered colonies in *M. tuberculosis* strains from transformations with 100 ng of *groEL1* AES. (C) Southern blot analysis of the *M. tuberculosis groEL1* gene locus of DNA isolated from 16 recombinant strains made by recombineering with a PCR-generated *groEL1* AES in a pJV53 strain. The probe anneals to the downstream homology region of the *groEL1* AES; expected band sizes: 7.5 kbp wild type; 10.2 kbp mutant. DNA from wild type and Δ*groEL1 M. tuberculosis* strains (mutant constructed by specialized transduction [14]) were used as controls. (D) Southern blot of DNA isolated from recombinants: pJV53 or pLAM12 cultures grown either in succinate and induced with acetamide, or in OADC and washed into succ./acet. The AES was generated by *Pac*I digest or PCR.

3.3.7 Recombineering efficiently targets replicating plasmids.

Throughout these experiments, the linearized AESs were generated by plasmid digest, with the exception of the *groEL1* AES, which was typically generated by PCR amplification. In one case, the *groEL1* AES plasmid was digested with an enzyme that cut near the *E. coli* origin of replication (*oriE*), and a large increase in the number of recombinant colonies (10^4) was observed (Figure 21A,B). This was not due to the presence of non-homologous regions at the ends of the substrate (data not shown), nor was it dependent on the presence of homologous sequences (pYUB854; Figure 21B). Recombinant colonies were analyzed by PCR with primers that anneal in the AES homologies and thus produced two PCR products (wild type and mutant) if the AES has been incorporated at a locus other than *groEL1*. These two PCR products were seen at a low frequency with either *Afl*II- or NcoI-digested AESs (Figure 21C). Conversely, all colonies yielded two PCR products from transformations with *Pci*I-digested plasmid AESs targeting the *groEL1* or MSMEG4308 genes (pMsgroEL1KO and pMs4308KO; Figure 21C,D).

These observations are likely connected to the presence of a region encompassing the *oriE* that is homologous to the extrachromosomal pJV24 plasmid (Figure 21A). Digest in this region (with *Pci*I) results in a linear AES with significant lengths of homology to pJV24 on each end. However, removal of the backbone of the plasmid produces colonies that are all correctly targeted to the chromosomal *groEL1* locus instead of the plasmid (Figure 21E). Analysis of plasmids electroduced from *Pci*I-generated colonies showed that 64% of the plasmids (originally Kan^R only) were Kan^R/Hyg^R (Figure 21F), and restriction digests revealed the presence of additional DNA sequences in these plasmids (Figure 21G). These data clearly demonstrate that extrachromosomal plasmids can be targeted by dsDNA recombineering, and that the plasmid backbone of circular AES constructs must be removed if chromosomal targeting is desired [227].

Figure 21. Recombineering targets extrachromosomal plasmids efficiently.

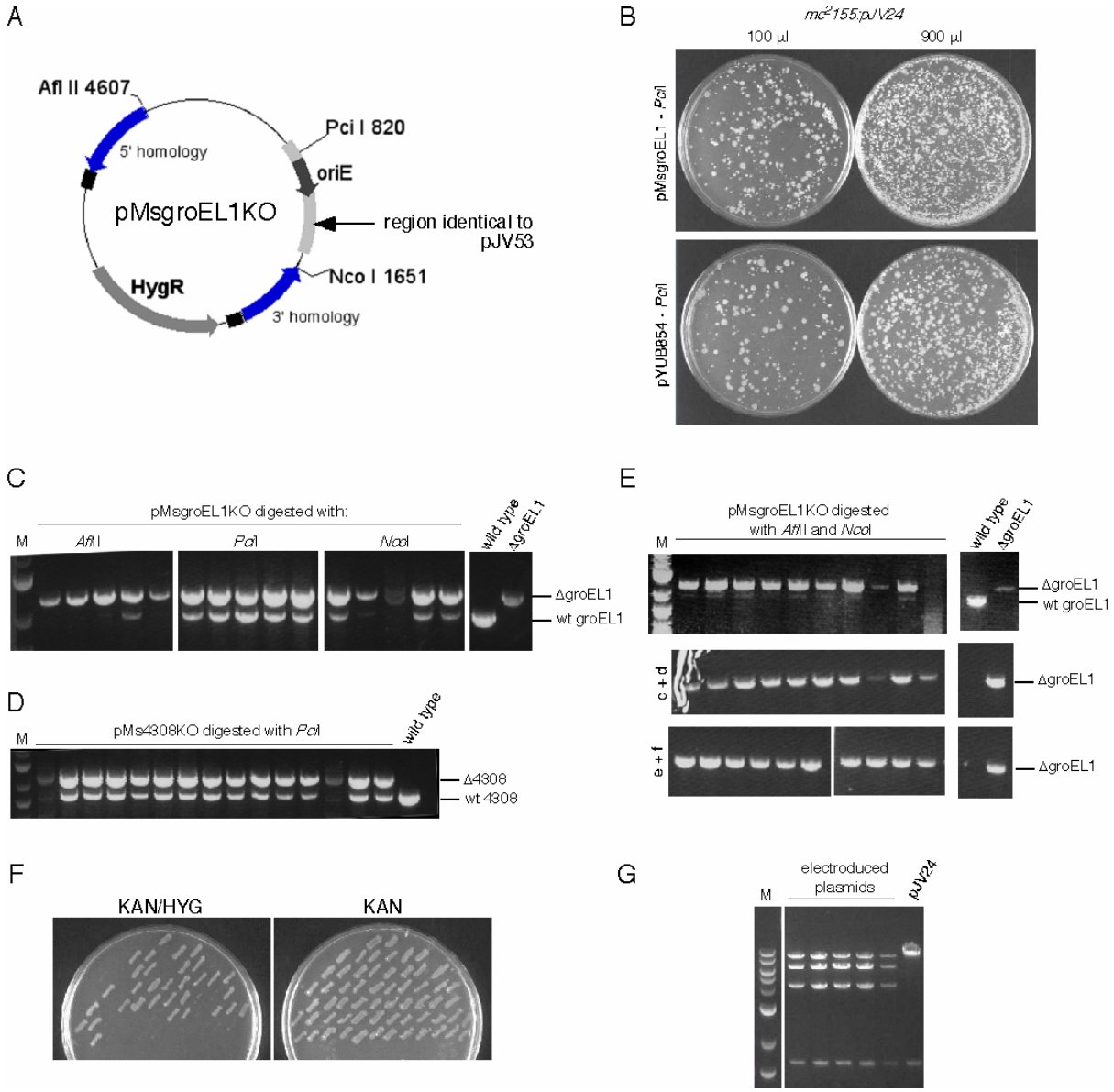


Figure 21. (A) The plasmid pMsgroEL1KO was constructed by cloning regions of homology to the *M. smegmatis groEL1* gene flanking a *hyg*^{R} cassette into the parent vector pYUB854 (not shown) [14]. These plasmids contain a region of homology to all mycobacterial extrachromosomal recombineering plasmids near the *E. coli* origin of replication (oriE), depicted by the light grey bar. (B) Transformation of 100 ng linearized pMsgroEL1 or the parent cloning vector without any homology to *groEL1* (pYUB854) digested with *Pci*I into mc^{2}155:pJV24 cells results in a large increase in recombinants. (C) Digestion of pMsgroEL1KO with either *Afl*II, *Pci*I, or *Nco*I results in two bands (mutant and wild type) by colony PCR with primers that anneal in the regions of homology. (D) Two bands are also seen in experiments targeting the *M. smegmatis* MSMEG4308 gene with an AES linearized by *Pci*I digest (pMs4308KO). (E) The correct bands are observed when pMsgroEL1 is double-digested with *Afl*II and *Nco*I, and all are correctly targeted using primers annealing either in (a+b) or outside (c+d, e+f) the homologous regions (see Figure 18 for primer locations). (F) Plasmids were electroduced from colonies (B) into *E. coli* and patched onto plates containing Kan or Kan/Hyg. (G) Restriction digests of five Kan^R/Hyg^R colonies shows multiple additional bands compared to pJV24 control.

The data presented in this chapter demonstrate that recombineering with dsDNA substrates is a simple and efficient method of constructing gene replacement mutants in both *M. smegmatis* and *M. tuberculosis* [227]. The success of this technique further suggested that other types of mutagenesis might be accomplished using recombineering, and these will be discussed in the following section.

3.4 POINT MUTAGENESIS

3.4.1 ssDNA recombineering of replicating plasmids requires only Che9c gp61.

To determine if short ssDNA substrates (oligonucleotides) could be used to make point mutations on mycobacterial genomes, a simple assay was developed using extrachromosomally-replicating plasmids as targets for recombination [228]. The chosen target gene was a mutated version of the *hyg*^{R} gene containing two consecutive amber mutations that inactivate its function (*hyg*^{S}); the assay therefore tests if a Hyg^R phenotype could be restored by ssDNA recombineering at the mutated locus. The *hyg*^{S} gene was cloned into various plasmids expressing Che9c gp60, gp61, or both (Figure 22A, Table 7), and electrocompetent cells were prepared of *M. smegmatis* strains containing these plasmids. Complementary substrates were synthesized that were 100 nt long (Table 16; JCV198, JCV199) and were homologous to the mutated region of *hyg*^{S}, with the mutations that restore wild type sequence in the center. Transformation of *M. smegmatis* strains expressing gp61 or both gp60/gp61 with either of these oligonucleotides resulted in more than 10^3 Hyg^R colonies, whereas strains expressing only gp60 or containing an empty vector had only background numbers of colonies (<25) (Figure 22B, Table 7). Transformations with

oligonucleotide JCV199 consistently resulted in higher recombineering frequencies, which may result from a strand bias due to the direction of DNA replication on this plasmid or a sequence-specific effect. Similar results were observed in *M. tuberculosis* strains expressing gp61 using this same assay, although at frequencies approximately 10-fold lower (Figure 22B, Table 7). Recovering transformations for three days (compared to one or two days) yielded the highest numbers of recombinants (data not shown). These data suggest that Che9c gp61 is sufficient for recombination with ssDNA substrates, and the number of recombinants generated by this method is 100- to 1000-fold greater than for what is obtained with dsDNA substrates.

Table 7. ssDNA recombineering of plasmids in *M. smegmatis* and *M. tuberculosis.*

	Strain background, hyg^S [a]	Recombinants recovered (Hyg^R)[c]		Recombineering frequency[d]		
		JCV198[b]	JCV199[b]	JCV198	JCV199	Ratio[e]
M. smegmatis mc²155	pJV73amber (control, P_{hsp60})	6	6	1.7×10^{-5}	1.7×10^{-5}	N/A
	pJV74amber (control, $P_{acetamidase}$)	23	0	7.1×10^{-6}	0	N/A
	pJV75amber (gp61, $P_{acetamidase}$)	5,300	3,310	3.9×10^{-2}	2.4×10^{-2}	1.6
	pJV76amber (gp60/gp61, $P_{acetamidase}$)	294,000	67,000	2.9×10^{-2}	6.6×10^{-3}	4.3
	pJV77amber (gp60, $P_{acetamidase}$)	1	0	8.0×10^{-6}	0	N/A
	pJV78amber (gp61, P_{hsp60})	2,710	250	3.5×10^{-3}	3.2×10^{-4}	10.8
M. tuberculosis H37Rv	pJV74amber (control, $P_{acetamidase}$)	3	3	9.4×10^{-7}	9.4×10^{-7}	N/A
	pJV75amber (gp61, $P_{acetamidase}$)	10,200	1,960	3.6×10^{-3}	6.9×10^{-4}	5.2
	pJV76amber (gp60/gp61, $P_{acetamidase}$)	2,130	1,020	1.3×10^{-3}	6.1×10^{-4}	2.1

a. Each plasmid (extrachromosomally-replicating; in strains of *M. smegmatis* or *M. tuberculosis* as indicated) contains a hyg^S gene with two codons mutated to early amber stop codons and various combinations of Che9c genes *60* and *61* under control of either an inducible promoter ($P_{acetamidase}$) or constitutive promoter (P_{hsp60}).
b. JCV198 and JCV199 are ssDNA oligonucleotides (100 nt; listed in Table 16) that are complementary, correspond to the mutated locus of hyg^S, and contain wild type sequence.
c. Number of Hyg^R recombinants with 100 ng of either JCV198 or JCV199.
d. Recombineering frequency is expressed as the number of recombinants per 100 ng ssDNA divided by the cell competency (expressed in cfu/μg DNA).
e. The ratio for each strain is calculated by dividing the recombineering frequency obtained from the ssDNA with the highest recombineering frequency by the other ssDNA. N/A: not applicable – background levels of recombinants.

Figure 22. ssDNA recombineering of plasmids in *M. smegmatis* and *M. tuberculosis*s.

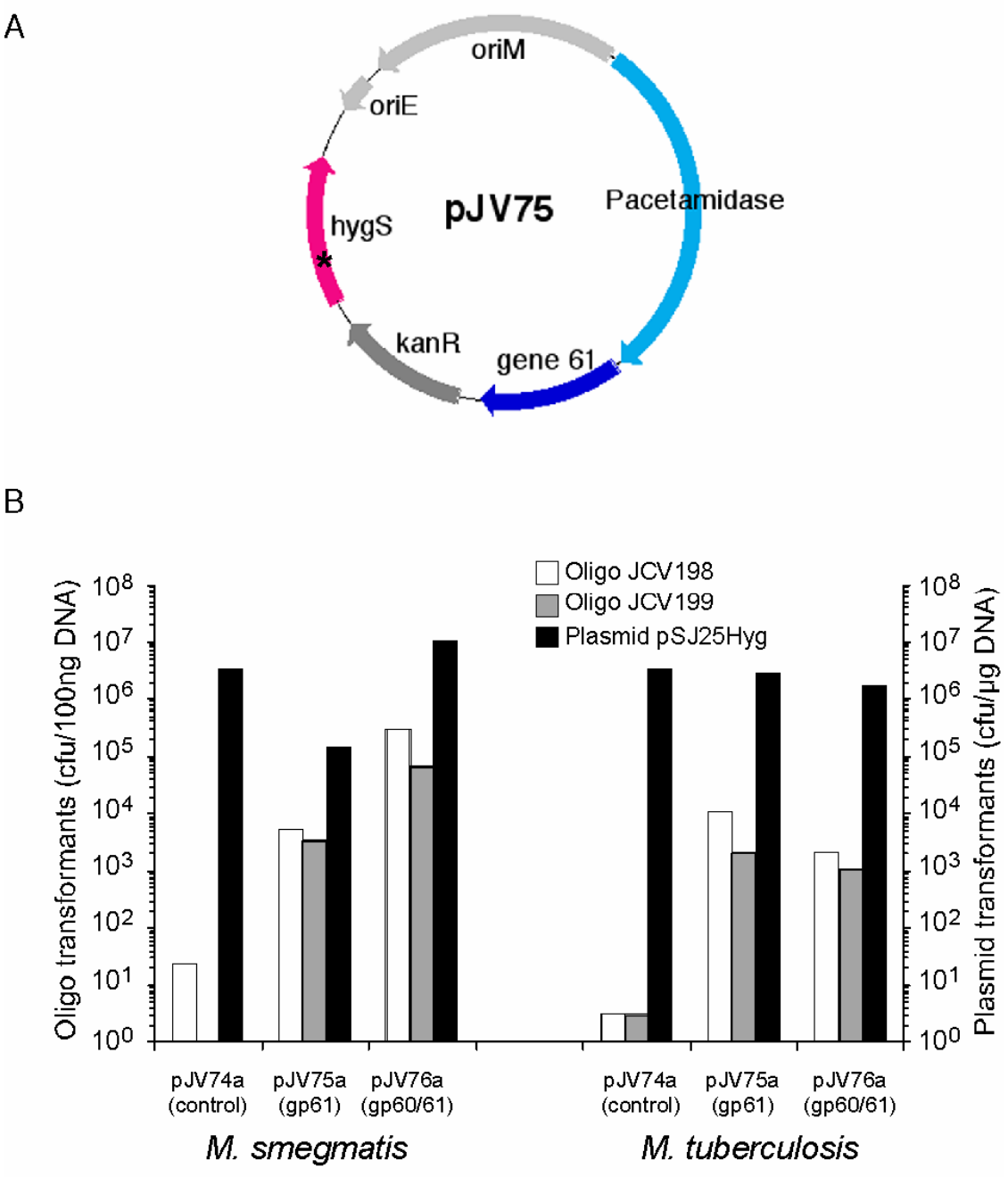


Figure 22 [228]. (A) Schematic of plasmid pJV75amber, an example of the plasmids constructed containing the *hyg*^{S} gene. (B) The number of Hyg^{R} transformants with 100 ng of JCV198 (white bars) and JCV199 (grey bars) (100 nt, complementary) are reported (left y-axis) for either *M. smegmatis* or *M. tuberculosis* strains containing plasmids pJV74amber (control), pJV75amber (expressing gp61 only), and pJV76amber (expressing both gp60/gp61). Cell competency is reported for each strain (black bars; right y-axis) as determined from control transformations with 50 ng of plasmid pSJ25Hyg and reported as transformants per µg plasmid DNA.

3.4.2 Introducing point mutations in the *M. smegmatis* chromosome by ssDNA recombineering

To determine if ssDNA recombineering could be used to target the chromosome, the same *hyg*^{S} gene was inserted into the chromosome of *M. smegmatis* at two loci using L5 and Bxb1 integration-proficient vectors [95,110]. The L5 and Bxb1 *attB* sites are located on different sides of the *M. smegmatis* chromosome and are approximately the same distance from the origin of replication (Figure 23A); Bxb1 is located 1.67 megabasepairs (mbp) 3′ of the origin, and L5 is 2.22 mbp 5′ of the origin (at 4.76 mbp). The *hyg*^{S} cassette was also inserted in both orientations at each locus in order to examine the possible strand bias seen with plasmid targeting (Figure 23B). Using these four strains, the same oligonucleotides (JCV198 and JCV199) were tested for ssDNA recombineering in an *M. smegmatis* mc^{2}155:pJV62 background, which expresses Che9c gp61 with transcription and translation signals from the acetamidase cassette. Hyg^R colonies were recovered using either oligonucleotide, although a strand bias was observed that, in some cases, was more than 1000-fold (Figure 23C, Table 8). This strand bias correlates with the direction of DNA replication at each locus, such that oligonucleotides that anneal to the lagging strand consistently generated higher recombination frequencies than those annealing to the leading strand. The *dif* site for replication termination is predicted to be at 3.41 mbp on the *M. smegmatis* chromosome [74,75], and bi-directional replication from the origin to the terminus correlates with the data from this assay (Figure 23). Interestingly, for each locus, one orientation resulted in a much smaller strand bias (<3-fold; Table 8, pJV89amber and pJV94amber); lower numbers of transformants were obtained with the oligonucleotide annealing to the lagging strand (JCV198), and higher numbers of transformants were found with the leading strand

oligonucleotide (JCV199) as compared to the other orientation. This effect was consistent in all strain backgrounds tested (Table 9) but was only seen with these particular integrated targets for ssDNA recombineering. This may be due either to the presence of additional genetic elements on the integrating plasmid that interfere with recombination or to a sequence-specific effect.

Table 8. ssDNA recombineering of a *hyg*^{S} gene in the *M. smegmatis* chromosome.

Strain background[a] pJV62 (gp61), *hyg*^{S}	Recombinants recovered (Hyg^R)		Recombineering frequency[c]		
	JCV198[b]	JCV199[b]	JCV198	JCV199	Ratio[d]
pJV89amber	1,220	670	3.2×10^{-3}	1.7×10^{-3}	1.8
pJV91amber	5	20,400	8.5×10^{-6}	3.5×10^{-2}	4,080
pJV92amber	15	27,600	8.4×10^{-5}	1.6×10^{-1}	1,840
pJV94amber	1,760	700	4.7×10^{-3}	1.9×10^{-3}	2.5

a. Each strain contains a *hyg*^{S} gene integrated at either the Bxb1 *attB* locus (pJV89amber, pJV91amber) or the L5 *attB* locus (pJV92amber, pJV94amber).
b. JCV198 and JCV199 are ssDNA oligonucleotides (100 nt; listed in Table 16) that are complementary, correspond to the mutated locus of *hyg*^{S}, and contain wild type sequence.
c. Recombineering frequency is expressed as the number of recombinants per 100 ng ssDNA divided by the cell competency (expressed in cfu/μg DNA).
d. The ratio for each strain is calculated by dividing the recombineering frequency obtained from the ssDNA with the highest recombineering frequency by the other ssDNA.

Recombineering of the *hyg*^{S} gene at these two loci was compared in strain backgrounds containing plasmids expressing either Che9c gp60/gp61 (pJV53) or only gp61, either from its endogenous translation signals (pJV62) or those of the acetamidase cassette (pJV52). Recombineering frequencies were approximately 10-fold higher in a pJV62 strain than in a pJV52 strain (Table 9), and notably, pJV62 expresses slightly lower levels of gp61 compared to pJV52 (Figure 15B). Therefore, strains expressing Che9c gp61 from plasmid pJV62 were used in most ssDNA recombineering experiments.

Table 9. ssDNA recombineering frequencies of chromosomal mutations in strains expressing Che9c gp61.

Strain background[a]	Recomb. Freq.[c] pJV89amber[b]		Recomb. Freq.[c] pJV91amber[b]		Recomb. Freq.[c] pJV92amber[b]		Recomb. Freq.[c] pJV94amber[b]	
	JCV198	JCV199	JCV198	JCV199	JCV198	JCV199	JCV198	JCV199
mc^2155:pLAM12 (control strain)	0	0	ND	ND	0	0	0	0
mc^2155:pJV52 (gp61)	ND	ND	ND	ND	0	6.9×10^{-2}	4.1×10^{-4}	1.8×10^{-4}
mc^2155:pJV53 (gp60/gp61)	ND	ND	ND	ND	6.0×10^{-5}	1.9×10^{-2}	2.9×10^{-4}	7.7×10^{-5}
mc^2155:pJV62 (gp61*)	3.2×10^{-3}	1.7×10^{-3}	8.5×10^{-6}	3.5×10^{-2}	8.4×10^{-5}	1.6×10^{-1}	4.7×10^{-3}	1.9×10^{-3}

a. Each strain contains an extrachromosomally replicating plasmid expressing Che9c gp60/gp61, only gp61, or is an empty vector. Plasmid pJV62 (*) expresses gene *61* from translational signals encoded by the acetamidase cassette.
b. Recombineering frequency is expressed as the number of recombinants per 100 ng ssDNA divided by the cell competency (expressed in cfu/μg DNA).
c. Plasmids containing a hyg^S gene integrated at either the Bxb1 *attB* locus (pJV89amber, pJV91amber) or the L5 *attB* locus (pJV92amber, pJV94amber) were integrated into the indicated strain backgrounds (a).

3.4.3 Recombineering chromosomal mutations that confer antibiotic resistance

Recombineering with ssDNA was also tested for the ability to introduce point mutations that confer resistance to antibiotics in the *M. smegmatis* chromosome [228]. These experiments were utilized to further characterize the strand bias observed with the hyg^S and to determine the advantages of using ssDNA recombineering for assessing the effect of a particular point mutation on antibiotic-resistance. Four well-characterized mutations were chosen: *inhA* S94A [11], *rpsL* K43R [94,212], *rpoB* H442R [94], and *gyrA* A91V [187], which are expected to confer resistance to isoniazid and ethionamide (INH/Eth), streptomycin (Str), rifampicin (Rif), and ofloxacin (Ofx), respectively. Each of these genes is located on one side of the chromosome (Figure 23C). Complementary oligonucleotides were designed to construct these specific mutations (Table 16), and these were transformed into *M. smegmatis* mc^2155:pJV62 cells.

Recombinant drug-resistant colonies were recovered at similar frequencies to those observed with *hyg*^{S} experiments (Figure 23B, Table 10); background levels of drug-resistance were similar to those reported in previous studies [11,94,187]. Targeting the lagging strand was most efficient for each gene (~10^5 colonies), which is consistent with the data from *hyg*^{S} targeting experiments and implicates a role for DNA replication in ssDNA recombination. Interestingly, the strand biases varied in size; the gene most proximal to the origin of replication (*gyrA*) had a strand bias of 36,000-fold, whereas the gene closest to *dif* (*inhA*) had a bias of 5-fold. Overall numbers of recombinants from targeting the *rpsL* gene were 10-fold lower than for other loci, although the strand bias (7,800-fold) was intermittent between *gyrA* and *inhA*. The bias for *rpoB* could not be determined accurately because the background level of spontaneous RIF-resistance masked the level of recombineering with the leading strand oligonucleotide (Figure 23C, Table 10). In addition, not only was the strand bias small at *inhA*, but this did not result from a decrease in colonies from the oligonucleotide targeting the lagging strand. Rather, there was an increase in recombinants with a leading strand oligonucleotide at this locus as compared to the others.

Table 10. Recombineering point mutations that confer drug-resistance in mycobacteria.

	Gene target [a]	Mutation[b]	Recombineering Freq. (mc^2155:pLAM12)[c]		Recombineering Freq. (mc^2155:pJV62)[c]		Ratio
			leading strand oligo	lagging strand oligo	Leading strand oligo	lagging strand oligo	
M. smegmatis mc^2155	*gyrA* [MSMEG_0006] (JCV259, lead; JCV260, lag)	A91V	1.4×10^{-6}	4.0×10^{-6}	8.5×10^{-7}	3.1×10^{-2}	36,000
	rpoB [MSMEG_1367] (JCV253, lead; JCV254, lag)	H442R	1.5×10^{-4}	1.0×10^{-4}	5.7×10^{-5}	2.2×10^{-2}	382
	rpsL [MSMEG_1398] (JCV218§, lead; JCV219§, lag)	K43R	0	0	6.4×10^{-7}	5.0×10^{-3}	7,833
	inhA [MSMEG_3151] (JCV216§, lead; JCV217§, lag)	S94A	6.6×10^{-4}	1.5×10^{-3}	6.5×10^{-3}	3.2×10^{-2}	4.9
M. tuberculosis H37Rv	*rpoB* [Rv0667] (JCV325, lead; JCV326 lag)	H451R	ND	2.1×10^{-5}	ND	3.6×10^{-3}	ND
	rpoB [Rv0667] (JCV327, lead; JCV328, lag)	S456L	8.8×10^{-6}	1.0×10^{-5}	2.1×10^{-6}	1.0×10^{-3}	480
	rpsL [Rv0682] (JCV329, lead; JCV330, lag)	K43R	1.5×10^{-6}	7.4×10^{-7}	3.6×10^{-7}	3.5×10^{-3}	9,722
	katG [Rv1908c] (JCV324, lead; JCV269, lag)	H108*	1.5×10^{-3}	1.4×10^{-3}	2.2×10^{-3}	5.7×10^{-4}	0.3

a. Specific drug-resistance mutations in *M. smegmatis* [MSMEG_X] or *M. tuberculosis* [RvX] genes were introduced by transformation with 100 ng of oligonucleotides that anneal to either the leading strand (lead) or lagging strand (lag) and are either 71 nt or 101 nt (§) in length.
b. The specific mutation introduced by the oligonucleotide (oligo) is indicated; *, amber.
c. Recombineering frequency is determine by the number of drug-resistant transformants for either empty vector control strain (pLAM12) or strain expressing gp61 (pJV62) divided by the cell competency. ND; not determined.
d. Comparison of the oligonucleotides: lagging strand divided by leading strand.

Figure 23. ssDNA recombineering of the *M. smegmatis* chromosome.

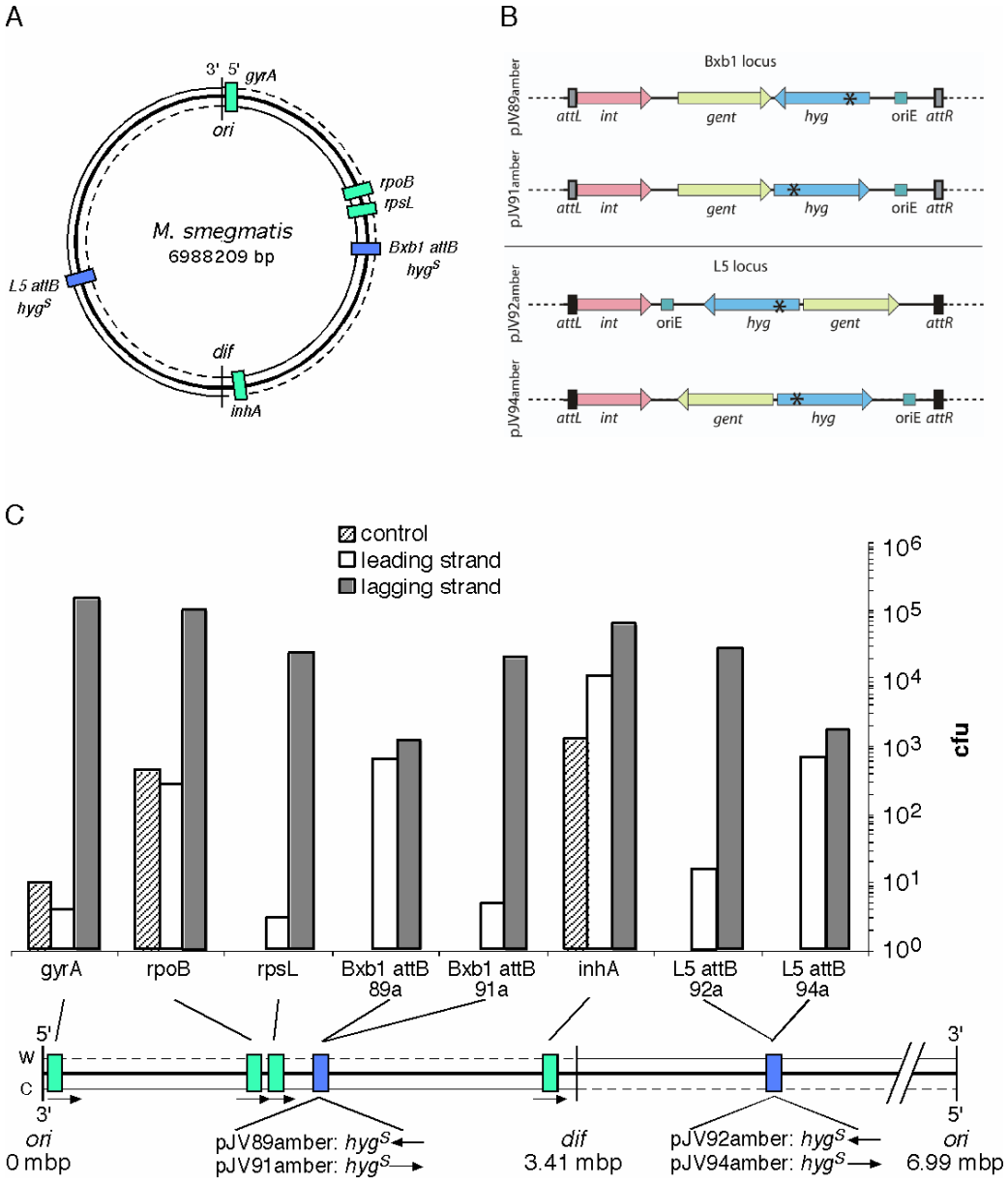


Figure 23. (A) Schematic of the location of genes targeted by recombineering on the *M. smegmatis* chromosome. The direction of DNA replication is predicted based on the location of the origin (*ori*) and terminus (*dif*) of DNA replication, which are indicated by solid (leading strand) and dashed (lagging strand) lines, as well as the size of the chromosome in mbp. The *hyg*^{S} gene is integrated at either the L5 or Bxb1 *attB* sites (blue); other gene targets are shown in green. (B) Illustration depicting the orientation of the *hyg*^{S} genes integrated at the L5 and Bxb1 loci. (C) The number of drug-resistant colonies obtained from transformations of a pJV62 strain (expressing Che9c gp61) with 100 ng of each oligonucleotide that anneal to either the leading strand (white bars) or lagging strand (grey bars) are shown in the graph (cfu). Background levels of spontaneous mutants for each drug are shown as determined from transformations of a control strain that does not express Che9c gp61 (hatched bars); background is zero for *hyg*^{S} and *rpsL*. The *M. smegmatis* chromosome is illustrated below the graph in a linear representation. The predicted strands for either leading strand synthesis (solid line) or lagging strand synthesis (dashed line) are shown. Arrows show the orientation of transcription for each gene. Drug-resistant colonies were selected with Ofx (*gyrA*), Str (*rpsL*), Rif (*rpoB*), INH/Eth (*inhA*), and Hyg (*hyg*^{S}). W: Watson strand; C: Crick strand.

Colonies from *inhA* targeting experiments were analyzed by PCR-amplification and sequencing of the *inhA* gene; all contained the S94A mutation, whereas colonies from negative control transformations did not (data not shown). An oligonucleotide that incorporates a synonymous third-base change at the same locus (*inhA* S94 codon) did not yield INH^R transformants above background levels. Recombination with ssDNA appears to be independent of host RecA (Table 11), unlike recombination with dsDNA substrates (Table 4). Collectively, these data indicate that introduction of the *inhA* S94A and other mutations arose from specifically-targeted recombination events that are dependent on Che9c gp61 and not general mutagenesis.

Table 11. ssDNA recombineering dependence on host RecA.

Strain background	Target: *inhA*[b]	Target: *rpsL*[b]	Recombineering frequency *inhA*	Recombineering frequency *rpsL*
mc^2155:pLAM12 (control)	1,630	1	3.2×10^{-4}	2.0×10^{-7}
mc^2155:pJV62 (gp61)	115,000	6,600	1.4×10^{-1}	8.1×10^{-3}
mc^2155:pJV62 *ΔrecA*[a] (gp61)	362,000	29,800	9.1×10^{-2}	7.5×10^{-3}

a. This *M. smegmatis ΔrecA* strain was constructed by K.G. Papavinasasundaram and colleagues [155] and is Hyg^R.
b. Oligonucleotides (100 ng) targeting the lagging strand containing either the *inhA* S94A (JCV217) or *rpsL* K43R (JCV219) point mutations were transformed into the *M. smegmatis* strains listed and either INH^R or Str^R transformants were selected, respectively. The number of transformants and recombineering frequencies for each target are shown.

Similar results were obtained in *M. tuberculosis* in which ssDNAs were designed to introduce point mutations in *rpoB* (S456L and H451R), *rpsL* (K43R), and *katG* (H108amber; INH^R [191,209]) (Figure 24, Table 10). Drug-resistant colonies (up to 10^4) were obtained with oligonucleotides that anneal to the lagging strand, with large strand biases up to ~9,700-fold; the background of *katG* prevented an accurate comparison of leading and lagging strand efficiencies. However, the recombineering frequencies were 5- to 30-fold lower as compared to those observed in *M. smegmatis* (Table 10), consistent with the plasmid-targeting results.

Figure 24. ssDNA recombineering of the *M. tuberculosis* chromosome.

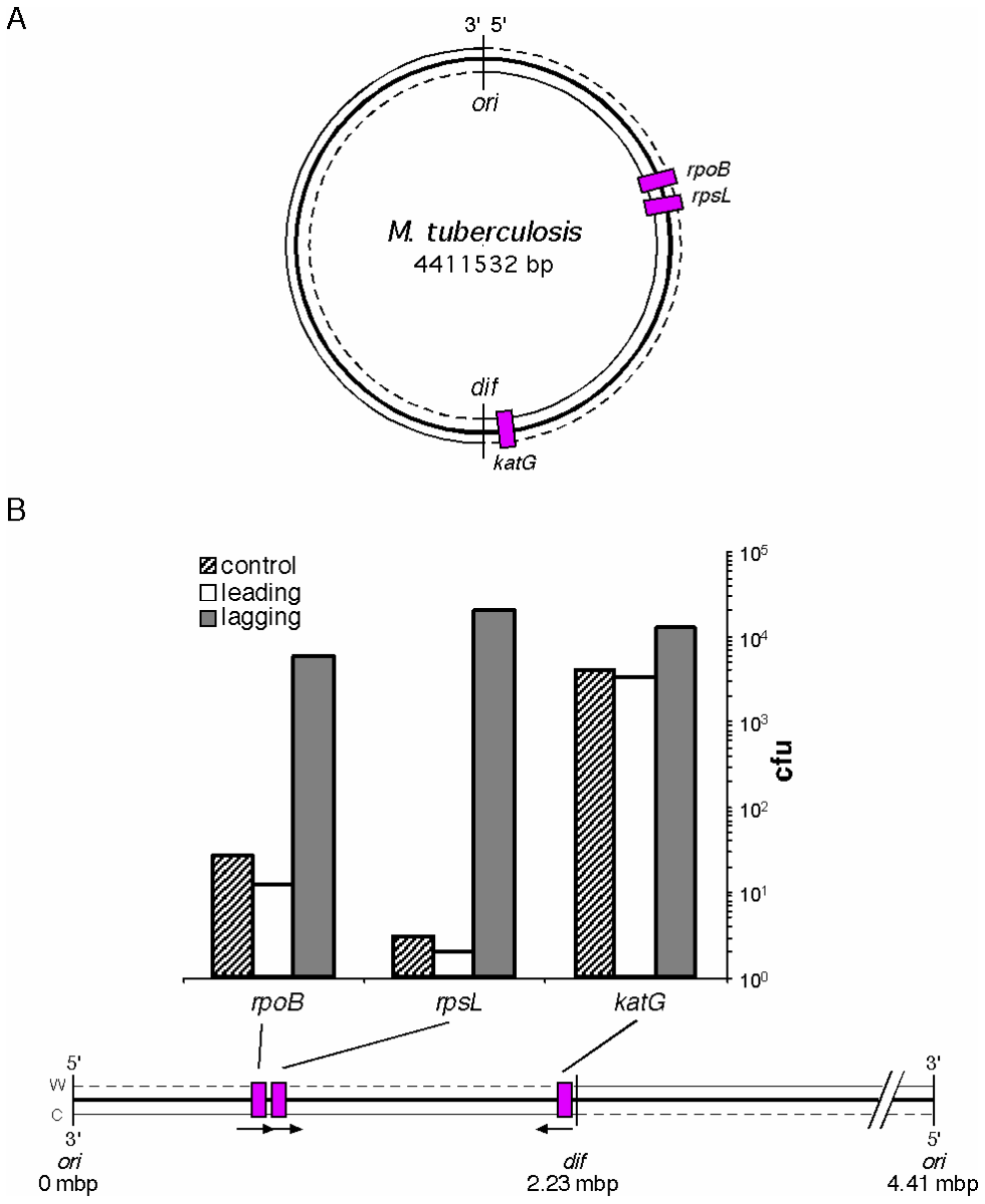


Figure 24. Similar to Figure 23 for the *M. smegmatis* chromosome, the locations of the *M. tuberculosis* genes targeted by ssDNA recombineering are depicted. (A) The location of the *rpsL*, *rpoB*, and *katG* genes on the *M. tuberculosis* chromosome, as well as *ori* and *dif* are shown on this schematic. Predicted leading (solid line) and lagging (dashed line) strands are indicated. (B) The number of drug-resistant colonies obtained from transformations of a pJV62 strain (expressing Che9c gp61) with 100 ng of each oligonucleotide that anneals to either the leading strand (white bars) or lagging strand (grey bars) are shown in the graph (cfu). Background levels of drug-resistant mutants are determined from transformations of a control strain that does not express Che9c gp61 (hatched bars). The *rpoB* mutation in this graph is S456L. Drug-resistant colonies were selected on Str (*rpsL*), RIF (*rpoB*), and INH (*katG*).

3.4.4 Optimizing ssDNA recombineering conditions

Several modifications to the ssDNA substrates were tested in order to optimize recombineering frequencies. First, to determine the effect of ssDNA length, the same assay described above using an integrated *hyg*^{S} gene was used [228]. Oligonucleotides of varying lengths (20 nt to 76 nt) were designed with the mutations that restore Hyg^R centrally located, and these were transformed into an *M. smegmatis* mc^{2}155:pJV62:pJV92amber strain. Maximal recombineering frequencies were achieved with oligonucleotides at lengths of 48 nt or greater, although low numbers of colonies were obtained above background at lengths as small as 32 nt (Figure 25). Since Che9c gp61 can bind a 20 nt oligonucleotide with similar affinity to a 44 nt oligonucleotide (Figure 12E), it is not clear why recombinants were not observed with oligonucleotides shorter than 32 nucleotides. The effect of ssDNA length is similar to what is observed with λ Beta, which works optimally with oligonucleotides 70 nt in length [52].

The effect of using dsDNA substrates was also examined at the *inhA* locus using 100 bp or 200 bp substrates (with a centered S94A mutation) in *M. smegmatis* cells expressing Che9c gp60 and gp61. This did not improve recombineering frequencies, and only slight increases in recombineering were observed in a similar assay using dsDNA substrates in *E. coli* [244]. Co-transformation of a Str^R ssDNA substrate with a plasmid that consitutively expresses Che9c gp61 into wild type cells did not result in recombinant Str^R colonies (data not shown), suggesting that gp61 must be expressed in the cell prior to transformation with the oligonucleotide. Similarly, pre-incubation of the ssDNA with gp61 prior to transformation into wild type cells did not yield recombinant colonies, an observation also made of λ Beta in the *E. coli* system [38]. Finally, the length of induction is optimal since the number of recombinants recovered is greatly increased at three hours compared to cultures without induction (up to 5,000-fold; data not shown).

Figure 25. ssDNA recombineering dependence on oligonucleotide length.

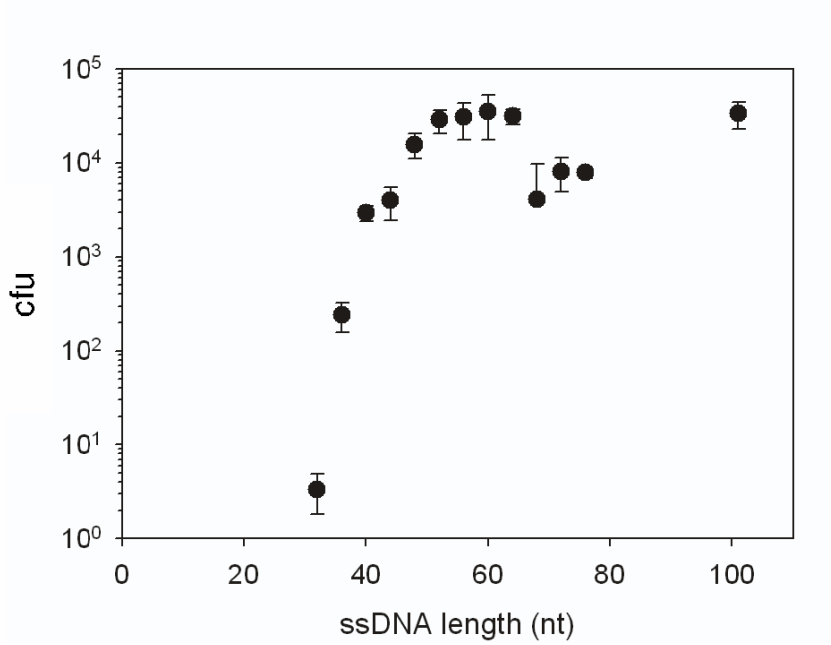


Figure 25 [228]. ssDNAs of varying lengths at four base intervals from 20 nt to 76 nt (and 100 nt as a positive control; JCV199) were tested for the ability to target the *hyg*^{S} gene integrated at the L5 *attB* site and restore Hyg^R. Lengths shorter than 32 nucleotides produced colonies at background levels. The error bars represent data from three independent experiments.

3.4.5 Development of a co-transformation strategy to select against non-transformable cells

The recombineering experiments performed with ssDNA substrates demonstrated that selectable point mutations could be made on either the chromosome or on extrachromosomal plasmid in *M. smegmatis* and *M. tuberculosis* [228]. However, most point mutations are not selectable by drug-resistance or other phenotypes, and instead require genotypic analysis to identify the mutant allele. Although recombination of ssDNA substrates is very efficient, approximately only one point mutant is recovered out of 1,000 viable cells. However, this frequency is similar to that observed in a standard plasmid transformation. Since the limiting factor appeared to be the competency of the cells, and not the frequency of recombination, it was reasoned that non-selectable point mutants could be recovered if the non-transformed cells could be removed from the population to be screened.

A co-transformation strategy was therefore tested in which a Hyg^R plasmid and an INH^R (*inhA* S94A) oligonucleotide were electroporated into *M. smegmatis* cells expressing Che9c gp61 and selected on media containing Hyg, INH, or both. Notably, INH^R/Hyg^R mutants were identified from colonies selected only on Hyg at a ~10% frequency (Figure 26A). This frequency was obtained using saturating amounts (500 ng) of either the plasmid or oligonucleotide, and 100 ng of the other substrate (Figure 26B). Similar co-selection frequencies were also obtained regardless of the type of Hyg^R plasmid, either integrating (Bxb1, L5, Giles) or replicating. This tactic was also successful using a double-oligonucleotide transformation: one oligonucleotide to introduce the desired point mutation (e.g. *inhA* S94A; JCV217; INH^R) and the other to repair a hyg^S mutation (JCV198) present on the extrachromosomal plasmid (pJV75amber). This resulted in a slightly lower co-selection frequency (~3-5%) but did not require the introduction of an

additional plasmid Optimal levels of co-selection were obtained with 200-500 ng of the INH^R oligonucleotide and 50-100 ng of the Hyg^R oligonucleotide (Figure 26C); increasing the Hyg^R oligonucleotide to 500 ng dropped frequencies ~10-fold. Finally, four hours recovery yielded optimal co-selection frequencies (7.2%), whereas shortening the time (1 hour, 1.2%) or lengthening the time more than 8 hours (overnight, 2.8%) did not improve recovery of doubly-resistant INH/Hyg colonies.

Figure 26. Optimizing recovery of point mutations by co-transformation of a Hyg^R substrate.

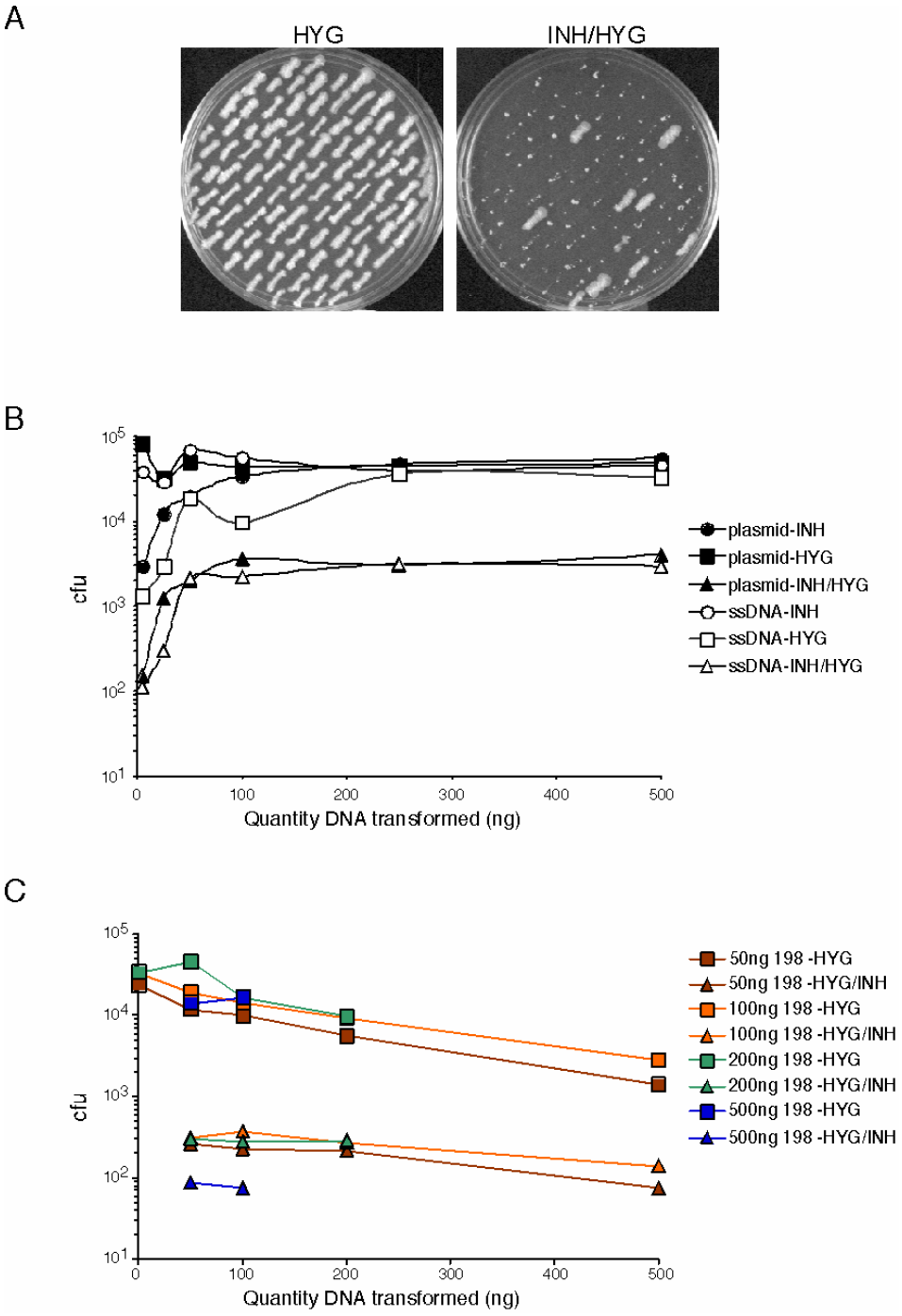


Figure 26 [228]**.** (A) Colonies from transformations with a Hyg^R plasmid (pSJ25Hyg) and an INH^R oligonucleotide (JCV217) were selected only on Hyg and patched onto INH/Hyg media. (B) Varying amounts of the Hyg^R plasmid and the INH^R oligonucleotide (10, 25, 50, 100, 250, or 500 ng) were co-transformed and plated on Hyg, INH, or Hyg/INH. The key indicates the substrate held constant at 500 ng (while the other substrate quantity was varied) and the type of antibiotic selection for each reaction. (C) Co-transformations with varying amounts of the Hyg^R oligonucleotide (JCV198) and INH^R oligonucleotide (JCV217) were plated on Hyg, INH, or Hyg/INH. The key indicates the substrate and quantity held constant, and the antibiotic selection for each reaction.

3.4.6 Point mutagenesis in the absence of selection

The results of the co-selection experiments suggested that mutant alleles could be easily identified at a high frequency when selection for plasmid transformants eliminates the non-transformable majority of the cell population. To test this idea, the same experiment was performed in which the INH^R oligonucleotide and Hyg^R plasmid were co-transformed, but the *inhA* mutation was not selected. Rather, the cells were diluted multiple times in liquid media containing Hyg in a culture block. Each culture well was plated to determine the number of starting cells; in this experiment, each well contained ~70 Hyg^R cells at the time of dilution. Culture wells were then grown to saturation, and the *inhA* locus examined for each well by mismatch amplification mutation assay PCR (MAMA-PCR) [30,219]. Mutant alleles (even a single base change) are identified in a large wild type population by this technique (Figure 27A), in which the primers and PCR conditions are optimized such that only mutant alleles are amplified. Using this method, at least one mutant *inhA* allele was identified in each culture well (Figure 27B). Homogenous mutant colonies were identified at a frequency of 3-4% from these culture wells by plating for single colonies and selecting for INH^R (Figure 27C). It is noted that this is slightly decreased from experiments in which colonies are selected directly following transformation on solid media (5-10%). However, this is likely due to variation between experiments, since Hyg/INH^R colonies were recovered at similar frequencies before and after outgrowth in culture wells in a different experiment (5% and 4.7%, respectively).

This technique was also tested with an additional gene locus, the *blaS* gene in *M. smegmatis*. The oligonucleotide, which was designed to introduce two consecutive amber mutations in *blaS*, was co-transformed with a Hyg^R plasmid, and the cells were diluted into liquid Hyg media as described above. MAMA-PCR identified a mutant allele in each culture

well (out of ~50 Hyg^R cells) (Figure 27D), and subsequent MAMA-PCR analysis identified pure mutant strains that arose from the plating of a positive culture well (Figure 27E). This experiment corroborated the previous experiments with *inhA*, in that mutant cells were present at ~3% out of Hyg^R cells. Only two rounds of PCR were required to identify three *blaS* mutants (43 PCR reactions). Alternatively, the experiment could likely be altered such that the transformed cells are diluted in Hyg media, grown to saturation, and plated for single colonies for MAMA-PCR analysis. Interestingly, the early amber mutations did not confer the expected ampicillin-sensitive phenotype for any of the three strains constructed; the *blaS* locus was analyzed by PCR and sequencing and the correct mutations were present (data not shown).

In addition, an *M. smegmatis pyrF* point mutant that introduces an early stop codon (Q61amber) was also constructed using co-transformation. Since inactivation of *pyrF* results in 5-FOA resistance and uracil auxotrophy, this mutant strain could also be used for co-selection in which uracil prototrophy acts as a positive selection, confirmed by 5-FOA sensitivity. The strain was constructed by double-oligonucleotide co-transformation, and three mutants were identified as Ura^S and $5\text{-}FOA^R$ out of 100 Hyg^R colonies. Two of these were confirmed by PCR and sequencing of the *pyrF* gene (data not shown).

Experiments targeting two additional non-selectable loci (*M. smegmatis groEL1* and *leuD*) were also attempted but were unsuccessful (data not shown). Several modifications were tested, including using oligonucleotides targeting both DNA strands, dsDNA substrates (100 and 200 bp), increasing the recovery time of the transformation or the amount of substrate, but no mutants were found by MAMA-PCR in any of the examined samples. A similar result was found during attempts to make a selectable point mutation in *M. smegmatis embB* (I289M) that should confer ethambutol resistance [112,214], although recombinant colonies were never obtained

above background levels. This may be due to sequence-specific effects, and experiments introducing different mutations at the same or different nucleotides could be performed to examine this possibility.

Finally, this strategy was tested on the *M. tuberculosis leuD* gene using co-transformation with a Hyg^R plasmid, and mutants were identified by MAMA-PCR (Figure 27F). The co-selection frequency in *M. tuberculosis*, estimated at 0.5% - 1%, is lower than that observed for most *M. smegmatis* loci. However, it should be noted that co-selection frequencies for the *M. smegmatis rpsL* locus are ~0.3 – 0.5%, since overall numbers of recombinants are ~10-fold lower at this target. Therefore, the low frequencies at the *M. tuberculosis leuD* target might be due to generally lower co-transformation or recombination frequencies, or sequence-specific effects. These data jointly demonstrate that both selectable and non-selectable point mutations can be constructed in *M. smegmatis* and *M. tuberculosis*.

Figure 27. Construction of non-selectable point mutations.

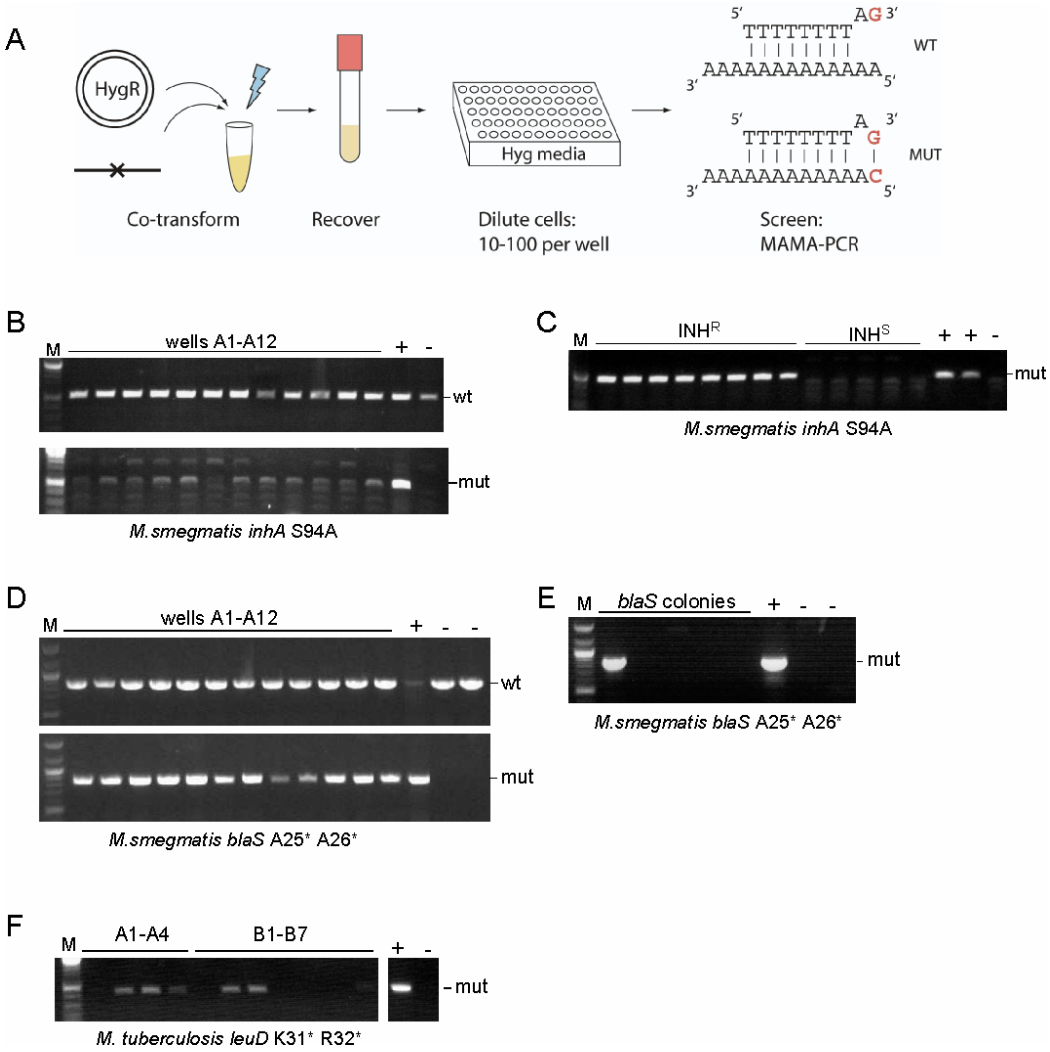


Figure 27. (A) Schematic of the strategy used to identify non-selectable point mutations by MAMA-PCR. Electrocompetent recombineering cells (expressing Che9c gp61) are co-transformed with a Hyg^R substrate (plasmid or ssDNA) and the ssDNA designed to introduce the desired mutation. Cells are recovered for four hours in media without antibiotics and diluted into culture wells (at multiple dilutions) in liquid media containing Hyg. MAMA-PCR primers are designed in which the penultimate base does not match either a wild type or mutant allele, but the ultimate base pairs only with the mutant allele. Using high fidelity *Taq* polymerase, PCR preferentially amplifies DNA from mutant alleles. (B-E) MAMA-PCR analyses using wild type or mutant primers of the *M. smegmatis inhA* and *blaS* loci after co-transformations. (F) MAMA-PCR analysis of the *M. tuberculosis leuD* locus after co-transformation. (B) Culture wells (12; A1-A12) from co-transformations with an oligonucleotide introducing an *inhA* S94A mutation and a Hyg^R plasmid show the presence of wild type (upper panel) or mutant (lower panel) alleles; positive (*inhA* S94A mutant) and negative (wild type DNA) controls are shown. (C) Analysis of the *inhA* locus of INH^R and INH^S single colonies isolated from culture well A1 show pure mutant alleles only for INH^R colonies. (D) Culture wells (12; A1-A12) from co-transformations with an oligonucleotide introducing two amber mutations (*) in *blaS* and a Hyg^R plasmid show the presence of wild type (upper panel) or mutant (lower panel) alleles; positive (mutant DNA made by PCR) and negative (wild type DNA) controls are shown. (E) Analysis of the *blaS* locus of single colonies isolated from culture well A1 shows a positive pure mutant allele. (F) Culture wells from co-transformations in *M. tuberculosis* with an oligonucleotide introducing two amber mutations (*) in *leuD* and a Hyg^R plasmid show the presence of mutant alleles. A1-A4: ~560 cells per well; B1-B7: ~56 cells per well.

3.5 OTHER APPLICATIONS OF RECOMBINEERING

An additional attractive use of recombineering is the construction of unmarked, in-frame deletion mutants, and this can also be used for removing antibiotic resistance markers from gene replacement mutants. To determine if mycobacterial recombineering could be used to make deletions, the *M. smegmatis leuD* gene was targeted with substrates that delete the same region as previous allelic replacement experiments at this locus (Figure 17A). However, in this case, short ssDNA (100 nt) or dsDNA (100 bp or 200 bp) substrates were used that did not contain an antibiotic marker for selection. In addition, dsDNA substrates were tested because previous experiments targeting mycobacteriophage genomes indicated that these substrates were more efficient for deletion construction than ssDNAs (L. Marinelli, manuscript in preparation). The co-transformation strategy was used for this experiment since it had been successful for constructing point mutations. Colonies were analyzed by two methods, either by plating directly following transformation recovery, or by diluting cells in liquid media containing Hyg. In one experiment, colonies plated directly following transformation (with a 100 bp *leuD* substrate and Hyg^R oligonucleotide) were replica plated onto media lacking leucine to identify leucine auxotrophs, and a single mutant was identified at a frequency of ~0.5% (Figure 28B).

In other experiments, culture wells were screened for the presence of the *leuD* deletion mutant. Mutants from experiments with an oligonucleotide that anneals to the lagging strand could be identified by MAMA-PCR at a low frequency (~0.2%), while no mutants were observed using a leading strand oligonucleotide (Figure 28C). Conversely, mutants were easily identified from transformations with 100 bp and 200 bp substrates using MAMA-PCR analysis, which indicated the presence of the mutant allele at a high frequency (Figure 28C). However, mutants can still be identified using a less sensitive PCR technique (with primers that anneal

outside the deletion locus). For example, in one experiment in which a Hyg^R plasmid was co-transformed with a 200 bp susbtrate, at least one mutant was identified out of eight culture wells (~10 cells per well; Figure 28D). Upon plating these colonies from the positive culture well, two mutants were identified and confirmed out of ten tested (Figure 28E). It should be noted that the frequencies observed by MAMA-PCR and standard PCR were inconsistent, which likely reflects the detection of much smaller quantities of the mutant allele by MAMA-PCR. Futher, mutant identification was simplified at this particular locus by screening for leucine auxotrophy. However, mutants were readily identified by using co-transformation and PCR screening techniques, and this technique is likely applicable to other genes. Additionally, it appears that using 200 bp substrates gives the highest recombineering frequencies, much like what is observed for mycobacteriophage recombineering (L. Marinelli, manuscript in preparation).

Figure 28. Construction of an *M. smegmatis leuD* unmarked deletion by recombineering.

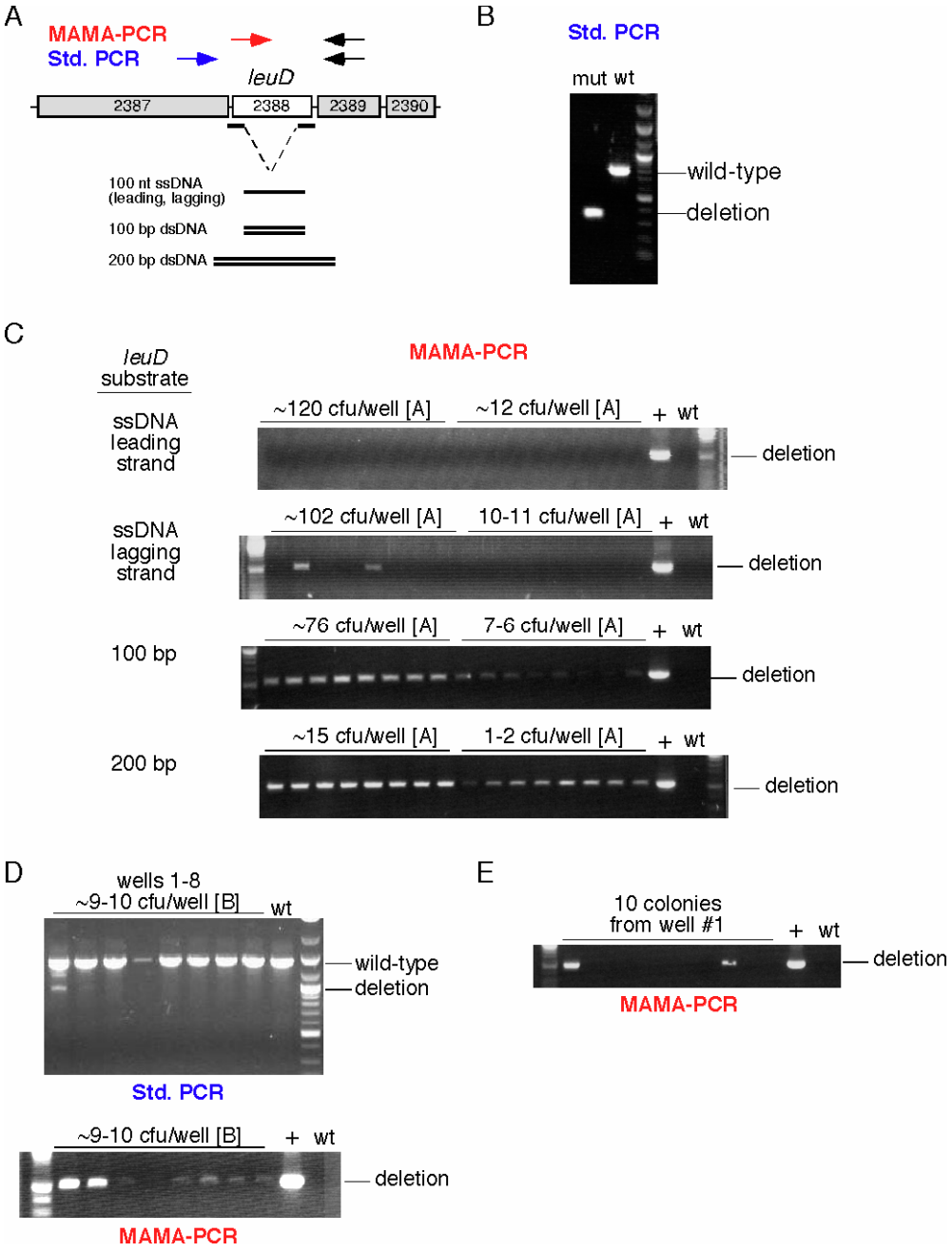


Figure 28. Recombineering of the *M. smegmatis leuD* gene to construct an unmarked deletion mutant. (A) A leucine auxotroph mutant was identified by replica plating following co-transformation with a 100 bp *leuD* deletion substrate and a Hyg^R oligonucleotide. Primers for standard PCR (Std.) anneal out side the targeted region (blue), whereas the MAMA-PCR primer anneals over the deletion junction (red). (B) A pure *leuD* mutant constructed with a 100 bp substrate and Hyg^R oligonucleotide selection is shown by standard PCR. (C) MAMA-PCR analysis of co-transformations (experiment [A]) with 100 nt ssDNA substrates (leading or lagging strand), a 100 bp and a 200 bp substrate, with a Hyg^R plasmid. The number of cells present in each well at the time of dilution (following transformation) are indicated. (D) Standard and MAMA-PCR analyses of culture wells from transformations (experiment [B]) with 200 bp substrate and a Hyg^R plasmid. (E) MAMA-PCR analysis of 10 single colonies from culture well #1 [B] of (D).

3.6 CONCLUSIONS

The Che9c recombineering system has successfully been used to construct gene replacements, point mutations, and gene deletions in the genomes of both *M. smegmatis* and *M. tuberculosis*. It is an efficient method for mutagenesis that is generally applicable to chromosomal and plasmid loci and is likely to be useful in other mycobacterial species. The Che9c proteins function similarly to the λ Red and RecET proteins in *E. coli* both *in vitro* and for recombineering *in vivo*. In fact, overall recombineering frequencies are similar between the two systems once differences in DNA uptake efficiencies are taken into account [52,240].

3.6.1 Recombineering: a powerful technique for constructing gene replacement mutants in the mycobacteria

Targeted gene knockouts can be made simply with linear AESs generated from circular plasmid constructs containing ~500 bp or more of homology to the target gene flanking an antibiotic resistance gene. These can be linearized either by PCR-amplification or double-restriction digest, ensuring removal of the plasmid backbone (Figure 29). Using 100 ng of these substrates generates a sufficient number of mutant colonies at every non-essential gene locus tested thus far, with more than 90% the desired mutants. Although gene knockouts were obtained using 50 bp of homology, this is not a recommended strategy due to the large decrease in recombineering frequency observed with these substrates. It is clear that the competency of the cells to take up DNA is an important criterion, and cells must be prepared with care. However, sufficiently competent cells were routinely made using a simple protocol without addition of glycine or other suggested supplements. As expected, the recovery of the desired gene replacement mutants is

dependent on expression of Che9c gp60 and gp61, and this sufficiently increased homologous recombination in *M. tuberculosis*, such that few recombinants arose from illegitimate recombination in these experiments.

Targeted gene replacement mutagenesis has obvious benefits for the potential of making large-scale ordered gene deletion mutants in the genomes of *M. tuberculosis* and *M. smegmatis*. Not only would this provide mutant strains for various experimental purposes, but it would also supplement the data regarding gene essentiality from previous genome-wide studies [200]. Additionally, nonsense mutations could be introduced into putative essential genes by ssDNA recombineering to confirm essentiality and to avoid the polar effects of gene replacements or transposon insertions.

Figure 29. Construction of a recombieering AES for allelic gene replacement mutagenesis.

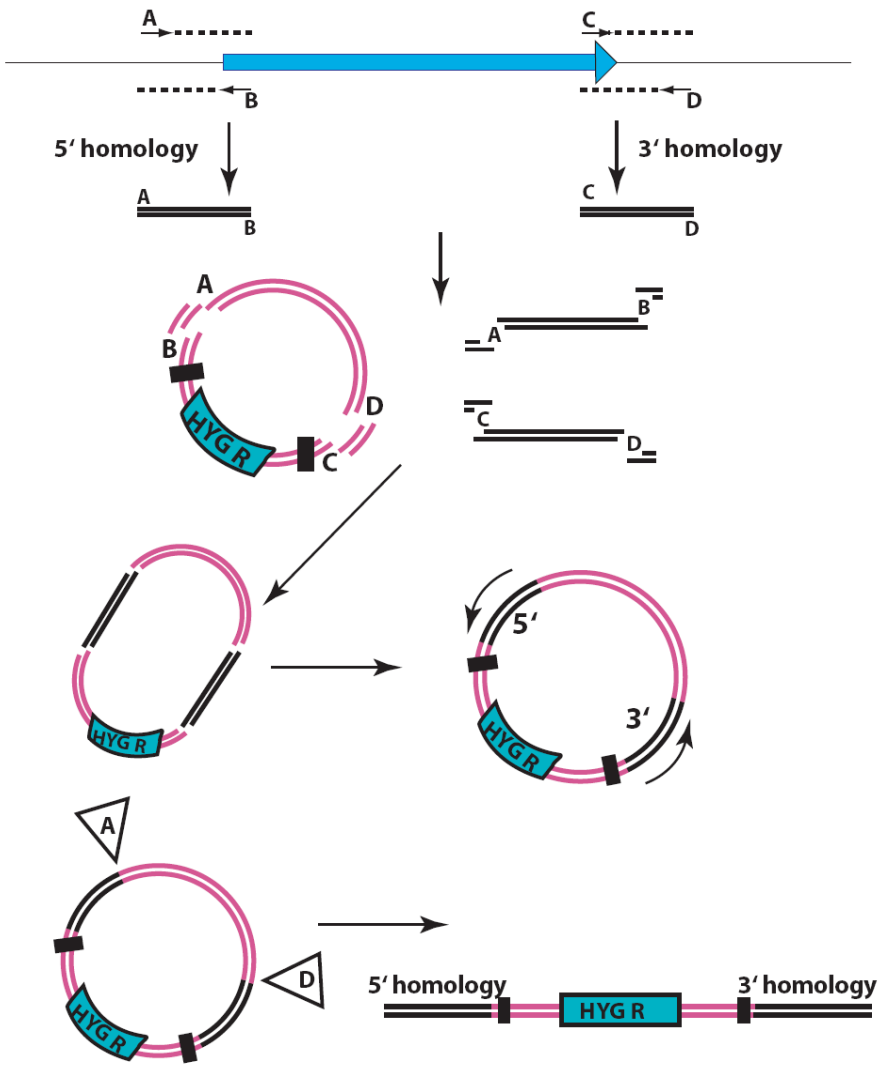


Figure 29. Diagram of the recommended procedure for generating recombineering AESs. The regions at the 5′ and 3′ ends of the targeted gene locus are amplified by PCR, such that the final products contain unique restriction sites (A, B, C, and D) for directional cloning flanking an antibiotic resistance gene (*e.g.* Hyg^R). The PCR products and the cloning vector are digested with all four enzymes and simultaneously ligated together. Selection of Hyg^R *E. coli* transformants generally yields a sufficient number of clones (albeit less than a standard two-way ligation), from which DNA is prepared and analyzed. Correctly cloned plasmids are digested with the two enzymes whose unique sites are at the distal ends of the homologous regions (A and D). Reactions are cleaned-up (removal of the plasmid backbone is unnecessary) and transformed into mycobacterial cells.

3.6.2 Recombineering of selectable and non-selectable point mutations

The mycobacterial ssDNA recombineering technology enables the construction of isogenic strains that differ by a single point mutation without direct selection, a technique that is unparalleled by any other approach. Point mutations can be introduced in mycobacterial chromosomes, replicating plasmids, and lytically-replicating mycobacteriophage genomes at high frequencies. Since only Che9c gp61 is required for ssDNA recombination, point mutants can be constructed without the potential toxic effects of expressing gp60. Because ssDNA substrates can be synthesized commercially, these experiments merely require design and purchase of ssDNAs of a minimum recommended length of 48 nucleotides, which eliminates the requirement for plasmid construction or other complex DNA manipulations. In addition, the double-oligonucleotide co-transformation strategy enables the construction of non-selectable point mutations, such that strains are completely unmarked following removal of the recombineering plasmid. Remarkably, the *E. coli* and mycobacterial recombineering systems perform comparably under optimal conditions for each system – MMR-defective, Gam-expressing strains and co-selective transformation, respectively – such that point mutants can be identified in the absence of direct selection at a high frequency (10-25%, respectively) [37,52,228].

The ssDNA recombineering technology could be specifically applied for determining the role of mutations that confer drug-resistance, particularly in regard to clinical research on the origins of XDR *M. tuberculosis* strains. Since most mutations are identified in combination with other mutations in *M. tuberculosis* strains, it is necessary to re-introduce each single mutation into a clean genetic background to determine its specific contribution to the strain's drug-susceptibility profile. However, this was previously not feasible due to the lack of generalized

transducing phages for *M. tuberculosis*. Only recently was the *inhA* S94A mutant strain constructed alongside an isogenic wild type strain by specialized transduction [232]. Similarly, the current study used ssDNA recombineering to test four characterized mutations for their ability to confer antibiotic resistance as reported, especially in genes such as *inhA* that have high levels of spontaneous mutagenesis and drug-resistance. For example, the *gyrA* A91V mutation had previously been identified *in vitro* along with other mutations in the *gyrA* gene, and this was not transduced into a clean genetic background [187]. The results of this study confirmed that this mutation is sufficient for generating resistance to ofloxacin. It is anticipated that similar experiments utilizing this technology will contribute to drug studies as a counterpart to detection and characterization of mutations conferring drug-resistance in *M. tuberculosis* clinical isolates.

3.6.3 Unique attributes of the mycobacterial recombineering system

Che9c gp61 functions similarly to λ Beta for ssDNA recombineering such that only the SSAP is required, recombination is independent of host RecA, and it is affected by the direction of DNA replication at the targeted locus. In both systems, ssDNA substrates that target (anneal to) the lagging strand DNA are more efficient than those that target the leading strand [52]. However, the strand biases in mycobacterial ssDNA recombineering assays at certain chromosomal loci are surprisingly quite sizeable (greater than 1000-fold in some cases), which is in stark contrast to the 2- to 50-fold bias in *E. coli* with λ Red recombineering [52]. It is clear that ssDNAs can recombine with the leading strand – since this was observed when targeting plasmids – but this is not consistently observed above background mutational frequencies or the limitations of DNA transformation. It is noted that experiments targeting plasmids in mycobacteria produce more modest strand biases, and this may be related to the mechanism of replication on the plasmid-

encoded origin of replication. Furthermore, the size of the strand bias decreased from the origin to the terminus, suggesting that regions near the replication terminus are replicated by forks traveling in both directions. Overall, the dissimilarity in strand bias magnitude observed in mycobacteria and *E. coli* may relect with fundamental differences in DNA replication, which is not well-characterized in the mycobacteria. Specifically, this could be related to availability of ssDNA regions near replication forks for ssDNA recombineering substrates, or alternatively, the interaction of the SSAP-bound ssDNA with host proteins involved in DNA replication.

Mycobacteria do not encode homologues of the *mutLS* MMR, nor do they appear to have a functional MMR system [213], which otherwise might influence the efficiency of generating certain basepair mismatches [113]. Thus, DNA replication appears to be the major factor contributing to the frequency of ssDNA recombineering in mycobacteria. However, transformation efficiency has a limiting effect on mycobacterial recombineering, and it is not clear if this masks any other contributing factors. For instance, there were several attempts to make chromosomal point mutations that were unsuccessful, and this may be due to undetermined sequence-specific effects at those loci. Also, it is not clear why the highest observed co-selection frequency is 10%. Although the frequency was not expected to surpass 50% since only one strand of the chromosome is targeted, it is surprising that it is as low as 5-10%, and the reason for this is not known.

An interesting difference between the *E. coli* and mycobacterial recombineering systems is the structure of the substrates required for making deletions. Experiments targeting both the mycobacterial chromosome and mycobacteriophage genomes for deletions demonstrated that dsDNA substrates yield better recombineering frequencies than ssDNA substrates, contrary to what is recommended in *E. coli* protocols for this type of mutagenesis [223]. In *E. coli*, ssDNA

substrates are clearly sufficient for making deletions at a high frequency [52,223], although the efficiency of dsDNA substrates for the same mutations has not been rigorously studied. Ultimately, 200 bp dsDNA substrates are recommended for constructing deletion mutants since these were more efficient than 100 bp substrates for both mycobacterial and mycobacteriophage genomes (see below). This is less suprising, however, since it reflects the correlation between an increase in homology length and an increase in recombineering frequency, which has been observed previously with both mycobacterial and *E. coli* recombineering systems [142,229,240].

3.6.4 Other uses for mycobacterial recombineering

An ideal use of recombineering is to unmark allelic gene replacement mutants, which would also be useful in the mycobacteria. This is accomplished in *E. coli* by using a ssDNA substrate with homology to the regions to be deleted, typically flanking an antibiotic resistance gene (used to select the gene knockout) and a SacB cassette for negative selection. Interestingly, as described above, short dsDNA substrates were more efficient than either ssDNA for making an unmarked deletion mutant in *M. smegmatis*. In-frame internal deletions and insertion of small tags for protein purification would also be useful strategies for assessing protein function. Although the ability to make these types of mutations in mycobacterial genomes has not been tested, it is likely that this will be successful, given the techniques that have been performed with the mycobacterial recombineering system on phages.

Experiments performed by Laura Marinelli have demonstrated that the mycobacterial recombineering system could also be used to target lytically-replicating mycobacteriophage genomes (manuscript in preparation). Phage mutagenesis is accomplished using a co-transformation strategy (similar to experiments described above) in which the phage DNA is

transformed into recombineering-proficient cells along with the substrate to make the mutation. Several types of mutations have been constructed in different phages, including point mutations, unmarked deletions, and small insertions, and even a mutation that introduced a deletion and point mutation simultaneously. The ability to make mutants in virtually any mycobacteriophage will facilitate the study of uncharacterized phage genes and systematic analysis of genes essential for phage propagation, among other uses.

3.6.5 Potential for optimizing the Che9c recombineering system

As with any newly developed system, there are a number of parameters that could be tested to further optimize the conditions and increase the number of recombinants recovered. It is clear that the level of protein expression plays a role in recombineering frequencies; however, higher levels of gp61 expression did not necessarily correlate with an increase in recombineering. This is observed in the *E. coli* system as well [135,137,244], and using a 5:1 ratio of Beta:Exo produces the best results (K. Murphy, personal communication). An approximation of this with the mycobacterial recombineering system was roughly attempted by placing Che9c *60* under control of the $P_{acetamidase}$ and *61* under the P_{hsp60}. This setup did not increase dsDNA recombineering frequencies, which is not surprising since constitutive expression of gp61 did not increase ssDNA recombineering. It is likely that testing other promoter combinations would produce better results. For example, using another inducible promoter such as the Tet operator/UV15 promoter system [51,67] to control expression might work well. Alternatively, the current induction conditions could be modified, such as by altering the concentration of acetamide.

Another attractive approach for potential improvement to the mycobacterial system is to determine if the λ Gam or another functional analogue (T4 gp2, P22 Abc1/2) might function in mycobacteria to block host nuclease degradation. However, due to the specific protein-protein interactions of Gam or Abc1/2 with RecBCD that are required for activity, these particular proteins may not be the ideal solution. Instead, it might be more beneficial to co-opt a system in which the proteins block degradation by a different mechanism, such as T4 gp2 or Mu Gam which bind and protect the ends of dsDNAs [2,6]. Alternatively, recombineering could be tested in *recBCD* mutant strains, although this limits the strain background. It is also not known if other mycobacterial nucleases act on dsDNA substrates (such as AddA), and therefore, mutation of host genes is not preferred. However, to ascertain the effect of nuclease mutations, as well as to begin to characterize the role of RecBCD and AddA in mycobacteria, some of these experiments have been performed and are described in the Appendix. Briefly, deletion of *recB* only modestly improved recombination frequencies (3- to 5-fold), and one assay with λ Gam did not improve recombineering frequencies. However, the Δ*recB* strain holds potential for certain recombineering methods in which the genetic background is of less importance, such as with recombineering of phage genomes. This strain could therefore improve the frequency of mutant allele recovery in these assays, making mutant isolation easier and more efficient.

Importantly, even though the Che9c recombineering system does not encode a Gam-like protein, it has worked sufficiently well for making gene replacement mutants at every non-essential locus thus far tested in the Hatfull lab (>20 genes). This is striking in that inhibition of RecBCD (either with Gam or using a Δ*recBCD* strain background) increases recombineering with the λ Red system at least 10-fold [43], and in some studies was found to be required [135,240,242]. This appears to be similar for RecET, such that Gam is not required but merely

enhances recombination efficiency [242]. It would therefore be interesting to examine the effect of Gam (or an analogous protein) on Che9c-mediated recombineering.

One concern is the potential for genetic rearrangements due to leaky expression of the Che9c genes in *M. tuberculosis*, when these cells are grown in succinate media without acetamide. This expression pattern has been observed previously in mycobacterial strains expressing proteins from this promoter cassette, but importantly, expression was repressed in media containing ADC [157]. Although Che9c gp60/gp61 expression has not been monitored in *M. tuberculosis* recombineering strains grown in ADC, it is likely that expression is decreased (if not repressed completely) in ADC media as compared to succinate media. Recombinants were obtained at similar frequencies when *M. tuberculosis* recombineering cells were grown in ADC, washed, and incubated in succinate/acetamide media, albeit the cell competency dropped for both strains tested. This therefore represents an alternative approach to possibly eliminate expression of the Che9c proteins prior to the required 24 hour induction.

Finally, other mycobacteriophage-encoded recombination systems were identified from sequencing and characterization. Although the Che9c proteins work sufficiently well for the types of mutagenesis thus far tested, other phages might encode recombination proteins with higher levels of activity. Therefore, the following chapter will examine the activity of phage-encoded recombination proteins in the mycobacteria.

4.0 IDENTIFICATION AND CHARACTERIZATION OF OTHER BACTERIOPHAGE RECOMBINASES

4.1 INTRODUCTION

Recombinases that function in the single strand annealing recombination pathway are found in many bacteriophages, although only a few have been well-studied. SSAPs are typically identified in operons adjacent to an exonuclease similarly to λ Red and *E. coli* RecET. These genes exhibit the mosaic pattern broadly observed in phage genomes, such that different genes encoding SSAPs and exonucleases are mixed [85]. For example, in phages SPP1 (of *B. subtilis*) and A118 (of *Listeria monocytogenes*), an Exo-like gene is found with a RecT-like gene. Several different exonucleases have been found in these systems, including proteins of the type II restriction enzyme fold (like λ Exo) and the type EndoVII fold [85]. Further, the gene order within the operon is not consistent, such that either the SSAP or the exonuclease can be transcribed first [43]. In some cases, like phage SPP1, other ORFs are predicted to lie between the genes encoding the exonuclease and recombinase [85]. Therefore, it appears that the organization of these recombination genes does not follow a particular pattern, but these can typically be identified based on similarity to other phage-encoded systems.

The apparent species-specificity of these recombination proteins is of particular interest with regard to the development and optimization of recombineering systems for bacteria other

than *E. coli*. A recent study by Datta and colleagues examined several putative SSAPs from phages that infect various bacterial hosts for activity in *E. coli* [43]. Using ssDNA recombineering as an assay, they observed that several SSAPs function with similar efficiency to λ Beta (~10^7 colonies), and not surprisingly, these proteins are predominantly from phages of Gram-negative bacteria. Several SSAPs from other phages that infect Gram-positive bacteria are able to introduce point mutations with moderate success (10^5 to 10^6 colonies). Notably, *B. subtilis* phage SPP1 gp35 and mycobacteriophage Che9c gp61 had the lowest recombination efficiencies (10^3 and 10^4 colonies, respectively). Also, a direct comparison of λ Beta and *E. coli* Rac prophage RecT showed that RecT functions ~30-fold worse in this assay. These data collectively suggest that there is a correlation between protein activity and organism relatedness, such that these phage-encoded proteins do not function as well in more distantly-related bacteria. Although the basis of this is unclear, it is possibly due to the ability of these proteins to interact specifically with host proteins during recombination, such as the components of the DNA replication machinery.

In the same study, it was also observed that the *B. subtilis* phage SPP1 gp34.1 and gp35 proteins, which are λ Exo and RecT homologues, respectively [85], promote dsDNA recombineering in *E. coli* [43]. Conversely, the *L. monocytogenes* phage A118 gp47 and gp48 proteins (also λ Exo and RecT homologues, respectively [85]) did not have dsDNA activity, although the gp47 had ssDNA recombineering activity [43]. Interestingly, the genes encoding SPP1 gp34.1 and gp35 are separated by three predicted ORFs, whereas A118 gp47 and gp48 are adjacent. These data suggest that different pairs of recombination proteins can be identified in phages, although they may not be located together and not all are necessarily active in this type of assay.

The finding that SSAPs from the same bacterial hosts – specifically, λ Beta and *E. coli* RecT – have different levels of activity has emphasized the need to examine other mycobacteriophage candidates for recombination activity. Mycobacteriophages are extremely diverse, and a large proportion of ORFs (~50%) do not have recognizable sequence similarity to known genes and are therefore of unknown function [165]. Currently, more than 50 mycobacteriophages have been sequenced, contributing to a vast reservoir of genetic information in which to search for SSAP-like genes. The Che9c-encoded recombinase and exonuclease were identified out of the first 14 sequenced mycobacteriophages. However, analysis of the more recently sequenced phages has revealed additional putative homologous recombination systems, and these will be discussed and characterized in this chapter.

Another approach to identifying phage-encoded recombinases is to assay directly for recombination activity, particularly with phages in which recombination proteins cannot be readily identified bioinformatically. During the construction of the first shuttle phasmids, Jacobs and colleagues observed that mycobacteriophage TM4 cosmid libraries recombined at a high frequency *in vivo* [86]. These TM4 cosmids are chimeric constructs in which an *E. coli* plasmid is randomly ligated to large fragments of the phage TM4 genome. These typically contain a nearly complete phage genomic molecule (containing an *E. coli* plasmid) with a small deletion of a portion of the phage DNA (see Figure 4). Most of these cosmids cannot be propagated individually as phages and are non-infectious (as assayed by plaque formation); however, those that can were further utilized as shuttle phasmids [13,14,86]. Strikingly, when plaques resulting from transformation with a pool of cosmids were analyzed, it was found that only one plaque out of 400 still contained the *E. coli* plasmid; the rest contained intact wild type TM4 genomic DNA. This suggested that TM4 encodes a recombination system, although none of the ORFs have

similarity to known recombination proteins [165]. Therefore, the same assay utilizing TM4 cosmid libraries was used to characterize the putative recombination system of this phage.

4.2 BIOINFORMATIC ANALYSIS OF OTHER MYCOBACTERIOPHAGE RECOMBINATION SYSTEMS

The results of BLAST analyses [4] using known recombination proteins as queries suggest that mycobacteriophages Halo, Giles, and BPs encode putative recombination systems that include recombinases in the λ Beta/RecT SSAP superfamily (Figure 30). Since Halo and BPs are 100% identical in the region containing these genes (99% overall), this analysis focused on the Halo proteins. Halo gp42 is 46% identical to Che9c gp60 and 30% identical to the C-terminus of RecE. However, analyses with Halo gp43 indicate that it is much more distantly related to other phage-encoded RecT proteins (~13% identity), and this was only identified after two rounds of PSI-BLAST. Additionally, the Halo gp43 protein was purified in a similar manner to Che9c gp61, and its DNA binding properties were analyzed by filter binding assays. Preliminary results indicated that gp43 does bind ssDNA (data not shown), although these experiments need to be repeated to determine the binding constant.

Unlike Halo, BPs, and Che9c, mycobacteriophage Giles does not have a RecE-like homologue; instead, gp52 contains a domain from the YqaJ family of phage-encoded exonucleases. The putative SSAP in Giles, gp53, is also not easily identifiable, but it has 30% amino acid identity to Halo gp43. These two proteins do appear to be members of the λ Beta/RecT SSAP family (Figure 30B), although more distantly related than Che9c gp61.

In the process of studying genes related to Che9c *60* and *61*, it was observed that *M. avium* contains a prophage that encodes similar proteins (Figure 30A). *M. avium* MAV_0829 shares 29% amino acid identity with Che9c gp61; it is annotated as a 'RecT/YqaK' protein and has 40% identity to *E. coli* RecT. Not surprisingly, the gene adjacent to this (MAV_0830) is predicted to encode an exonuclease; it is 41% identical to Che9c gp60 and 23% identical to RecE. Further BLAST analysis also identified an Erf-like protein, gp64, encoded by mycobacteriophages Wildcat and Cjw1; Wildcat gp64 is 21% identical to P22 Erf and Cjw1 gp70 is 15% identical to Erf (not shown). However, none of their adjacent genes has similarity to known recombination proteins. A prophage in *M. abscessus* is also predicted to encode a protein that is a distant relative of Erf (MAB_1744; 18% identity; not shown). Collectively, these data provided additional candidates to test for recombination activity in mycobacteria.

Figure 30. Mycobacteriophage-encoded recombination systems.

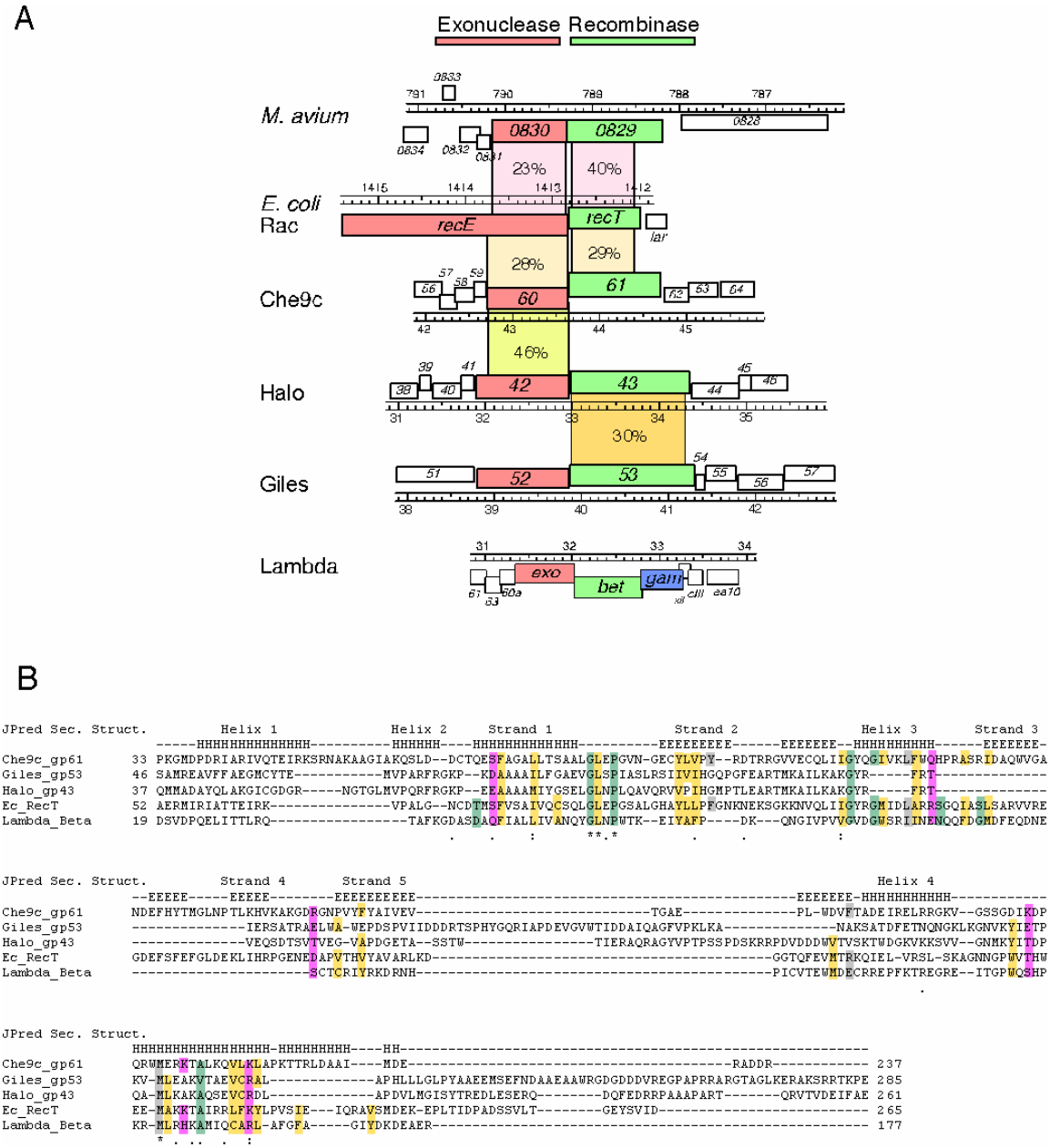


Figure 30. Mycobacteriophages Giles and Halo, and an *M. avium* prophage encode putative recombination systems. (A) Che9c gp60, Halo gp42, and *M. avium* MAV_0830 are RecE homologues, while Giles gp52 has identity to a domain from the YqaJ-like exonuclease family. Che9c gp61, Halo gp43, Giles gp53 and *M. avium* MAV_0829 are RecT homologues. Proteins that share more than 20% amino acid identity are connected by shaded boxes and percent identity indicated. Exonucleases are indicated in red, and SSAPs (recombinases) are indicated in green. *E. coli* Rac prophage genes, Halo genes, Giles genes, *M. avium* genes, and Che9c genes are transcribed from left to right, while the λ genes are transcribed right to left. (B) Multiple sequence alignments were constructed with Che9c gp61, Halo gp43, Giles gp53, λ Beta, *E. coli* RecT, *Shigella dysenteriae* Beta, and *Listeria innocua* Lin1755; the last two were removed following alignment for simplicity. The alignment was performed similarly to Iyer *et al.* [85] using T-coffee [148], and secondary structure predictions (using JPred) were also conserved [40]. Similar residues are highlighted that were found by Iyer *et al.* to be conserved greater than 85%.

4.3 COMPARISON OF SSAP ACTIVITY IN *M. SMEGMATIS*

Recombineering with ssDNA substrates provides a simple assay for recombination activity *in vivo*. In particular, drug-resistance mutations that give low background can be introduced in the *M. smegmatis* chromosome (discussed in section 3.4.3). Several SSAPs were tested for activity, including Halo gp43 and Giles gp53. In addition, *E. coli* RecT and λ Beta were analyzed in order to determine their activity in this distantly related bacterium, as well as to compare these results to those from a similar study by Datta *et al.* performed in *E. coli* [43]. The SSAP genes – Che9c *61*, Halo *43*, Giles *53*, *E. coli recT*, and λ *bet* – were under control of the $P_{acetamidase}$ promoter cassette (plasmid pLAM12) such that translation was derived from either their endogenous signals (pJV52, pJV103, pJV145, pJV104, and pJV105) or from signals in the acetamidase promoter cassette (*Nde*I site; see Figure 14) (pJV62, pJV106, pJV116, pJV107, and pJV108).

M. smegmatis strains containing each of these plasmids were induced for expression the same way as all previous recombineering strains with the Che9c proteins. Reverse-transcription PCR (RT-PCR) analysis was used to examine expression from $P_{acetamidase}$ for many of the constructs expressing genes from their endogenous translation signals (Figure 31A). Western blot analysis was performed on strains expressing Halo gp43 (Figure 31B), but was not suitable for some strains. Antibodies were not available for *E. coli* RecT, and the λ Beta antibodies (a gift from D. Court) had high background signal in *M. smegmatis* cells that masked any potential protein expression. Neither RT-PCR nor western blots were performed on Giles gp53 protein expressing strains. Strains expressing λ *bet* and *E. coli recT* both had detectable levels of RNA after induction with acetamide, indicating that at least transcription from this promoter is active in these constructs. However, this does not rule out any potential translation or protein instability

problems that could occur in this assay. Similar to Che9c gp61 (Figure 15), protein expression was also observed for the Halo gp43 (Figure 31B) by western blot.

Following these analyses, in order to test *in vivo* recombineering activity, each strain was transformed with oligonucleotides that introduced point mutations in the *inhA*, *rpsL*, and *gyrA* loci and recombinant colonies were selected. There were several surprising observations from these assays (Figure 31C, Table 12). First, Halo gp43 had a significant level of ssDNA recombineering activity, whereas Giles gp53 did not. This may be due to a lack of adequate expression, which can be tested in the future by RT-PCR and/or western blot analysis. Second, *E. coli* RecT had a high level of activity that was similar to Halo gp43, although not as high as Che9c gp61. Strains expressing λ Beta did not produce recombinant colonies above background, which is expected given anecdotal reports that the λ proteins are not active in mycobacteria. In addition, the strain that expressed RecT from the translation signals of the acetamidase promoter cassette had higher levels of activity (~10-fold) than the strain expressing RecT from its own translation signals (Figure 31C); this was also observed for Che9c gp61 in this and in previous assays (Table 9).

These data suggest that Che9c gp61 has the highest level of ssDNA recombineering activity in mycobacterial cells. Further, although Halo gp43 and Giles gp53 are 30% identical, strains expressing the Halo protein produced recombinants, whereas strains containing the Giles constructs did not. Finally, there is a substantial difference between the activities of *E. coli* phage SSAPs in the mycobacteria, such that λ Beta cannot recombine ssDNAs, while RecT is moderately efficient.

Table 12. Comparison of SSAP recombination activities in *M. smegmatis*.

	Strain (plasmid)	Protein	Cell Comp.[a]	*inhA*		*rpsL*		*gyrA*	
				cfu[b]	Rec. Freq.[c]	cfu[b]	Rec. Freq.[c]	cfu[b]	Rec. Freq.[c]
	pLAM12	[control]	2.7×10^{6}	ND	ND	4	1.5×10^{-6}	3	1.1×10^{-6}
Endogenous signals	pJV52	Che9c gp61	4.1×10^{5}	49,000	1.2×10^{-1}	2,200	5.4×10^{-3}	ND	ND
	pJV103	Halo gp43	6.6×10^{5}	10,100	1.5×10^{-2}	221	3.4×10^{-4}	ND	ND
	pJV145	Giles gp53	1.1×10^{6}	ND	ND	4	3.7×10^{-6}	9	8.3×10^{-6}
	pJV104	*E. coli* RecT	7.8×10^{5}	3,190	4.1×10^{-3}	101	1.3×10^{-4}	ND	ND
	pJV105	λ Beta	3.6×10^{5}	1,740	4.8×10^{-3}	2	5.6×10^{-6}	ND	ND
$P_{acetamidase}$ translation signals	pJV62	Che9c gp61	6.2×10^{5}	ND	ND	7,900	1.3×10^{-2}	54,000	8.7×10^{-2}
	pJV106	Halo gp43	5.6×10^{5}	ND	ND	218	3.9×10^{-4}	2,420	4.3×10^{-3}
	pJV116	Giles gp53	1.6×10^{5}	ND	ND	2	1.3×10^{-5}	23	1.5×10^{-4}
	pJV107	*E. coli* RecT	1.3×10^{6}	ND	ND	1,400	1.0×10^{-3}	9,700	7.2×10^{-3}
	pJV108	λ Beta	8.6×10^{5}	ND	ND	14	1.6×10^{-5}	1	1.2×10^{-6}

a. Cell competency is determined by transformation with 50 ng of a control plasmid; expressed in cfu/μg DNA.
b. The number of drug-resistant transformants using 100 ng oligonucleotide to introduce the following mutations: *inhA* S94A (INH^R), *rpsL* K43R (Str^R), or *gyrA* A91V (Ofx^R). ND; not determined.
c. Recombineering frequency is determined by dividing the number of recombinant colonies (b) by the cell competency (a). ND; not determined.

Figure 31. Comparison of SSAP recombination activities in *M. smegmatis*.

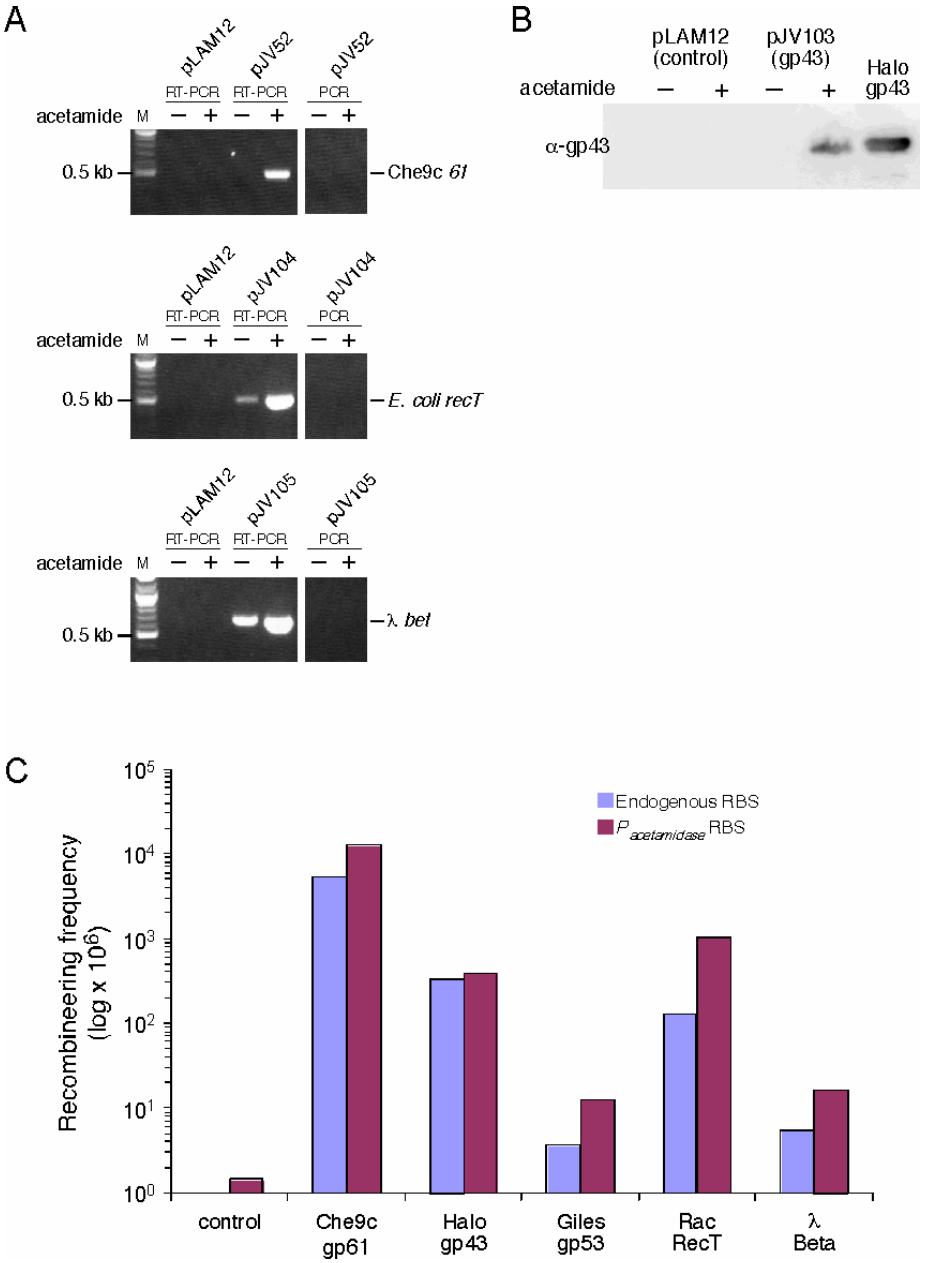


Figure 31. [228] (A) RT-PCR analysis of RNA extracted from *M. smegmatis* cultures in the presence or absence of induction with acetamide. RT-PCR products were analyzed with gene specific primers (Table 16) from strains containing the following plasmids: pLAM12 (empty vector), pJV52 (Che9c *61*), pJV104 (*E. coli recT*), and pJV105 (λ *bet*). Sizes of expected products: 482bp, 507 bp, and 642bp, respectively. PCR reactions without reverse transcriptase present were tested for the presence ofcontaminating DNA in the samples as a negative control. (B) Western blot analyses of strains expressing Halo gp43 in the presence or absence of inducer (0.2% acetamide) with polyclonal antibodies generated against purified gp43. (C) Recombineering frequencies of *M. smegmatis* strains expressing various SSAPs are shown from transformations with an oligonucleotide (JCV219) that confers Str^R (*rpsL* K43R). The frequencies are represented on a log scale, and the frequencies are multiplied by 10^6 for presentation purposes. *M. smegmatis* strains contain plasmids that express SSAP genes from either their endogenous translation signals (RBS, pLAM12 *Hpa*I site) or from translation signals present in the acetamidase promoter cassette (RBS, $P_{acetamidase}$; pLAM12 *Nde*I site); see Figure 14 for plasmid pLAM12 details.

4.4 CHARACTERIZATION OF A PUTATIVE RECOMBINATION SYSTEM IN MYCOBACTERIOPHAGE TM4

Although the number of sequenced genomes continues to increase, recent PSI-BLAST analysis of the predicted gene products encoded by mycobacteriophage TM4 still did not reveal any clues as to which genes might provide the recombination activity. While some gene products had short regions of similarity to proteins with known recombination activity, they were not good candidates or were not located next to genes that were likely to be part of a recombination system. For example, gp54 (93 amino acids) has similarity to a region of the YqaJ-like exonuclease protein of *Bacillus cereus*. However, gp53 has only low levels of sequence similarity with hypothetical transpeptidase or dehydrogenase proteins, whereas the other adjacent gene, gp56, is predicted to encode a protein only 29 amino acids in length with no sequence similarity to known proteins. In another case, gp59 has only 17% sequence identity to a putative RecB family exonuclease from a Thermus phage. Again, the adjacent genes are not likely candidates; gp57 is predicted to encode a DinG helicase, gp58 has an esterase_lipase domain, and gp60 is a small protein (57 amino acids) without similarity to any proteins in the database. It is possible that there are different start sites for some of these genes that would alter the analysis. However, based on the annotated genes, these do not appear to encode *bona fide* recombinase or exonuclease homologues.

Therefore, to further examine the recombination phenotype observed for the TM4 cosmid molecules, a new TM4 cosmid library was constructed (as described [86] and in Figure 4, except a Hyg^R plasmid was inserted instead of an Amp^R plasmid). These molecules contained the *E. coli* plasmid inserted in a region of the TM4 genome that is either essential (non-viable phage = cosmid) or non-essential (true shuttle phasmid). TM4 cosmid DNA was isolated from individual

E. coli Hyg^R colonies and examined by analytical restriction digest and sequencing to determine the structure of the cosmid. A set of cosmids was obtained by three rounds of screening, and the location of the *E. coli* plasmid and size of the TM4 deletion region were determined; these are illustrated in Figure 32. As a reference, four shuttle phasmids that were isolated by Jacobs and colleagues are also depicted in Figure 32 [13,14,86,87,210]. In addition, Hyg^R colonies (~27,000) were pooled, and DNA was prepared in order to repeat the experiments performed by Jacobs *et al.* [86].

Figure 32. Diagram of the TM4 cosmid library.

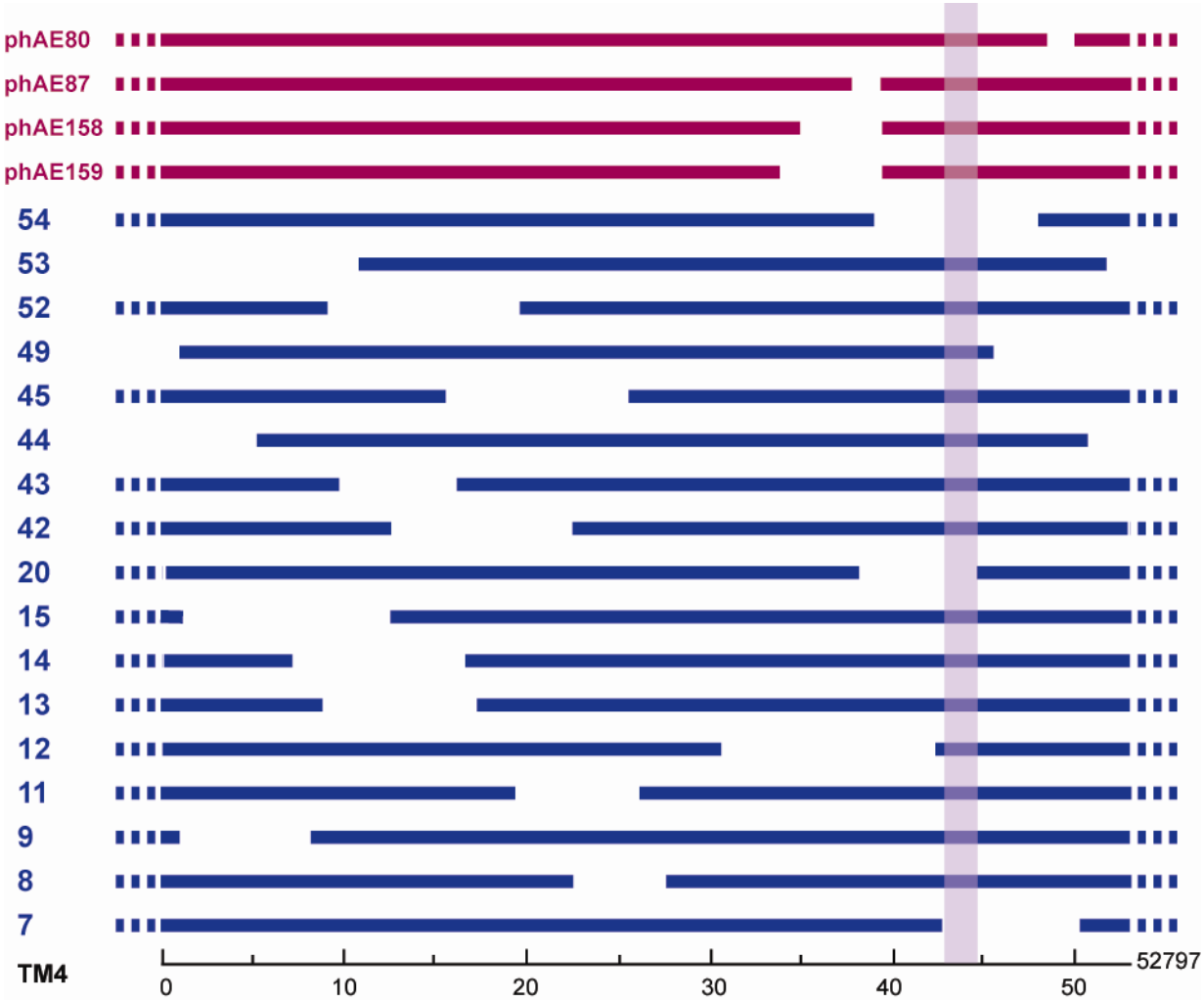


Figure 32. Each solid line on the schematic represents a TM4 cosmid (TM4cosX) or shuttle phasmid (phX). Dotted lines at the ends indicate that the molecule is connected at the termini; otherwise the *E. coli* plasmid connects the circle. The 'blank' spaces indicate the region deleted in the cosmid, as well as the location of the *E. coli* plasmid. A linear representation of the TM4 genome is depicted below (in kbp). Shuttle phasmids were made by Jacobs and colleagues [13,14,86]. The purple box indicates the region deleted in both TM4cos7 and TM4cos20 that renders them incapable of recombination.

Recombination experiments were performed in which cosmid pairs were co-transformed (500 μg each) into wild type *M. smegmatis*, recovered for 30 minutes, and plated as top agar lawns with additional *M. smegmatis* cells. Plaque numbers were recorded for each transformation. It was observed that only cosmid pairs that represented the full genome between them (*i.e.*, non-overlapping deletions) could produce plaques (Table 13), supporting the hypothesis that these molecules undergo recombination *in vivo*. Since most cosmids had large deletions of their genome (~9 kbp), it is not surprising that the individual cosmids could not propagate as phages. Pairs of cosmids with overlapping deletions did not produce plaques, likely because the common deleted region was essential. As expected, plaques were also obtained from transformations with DNA of the 'pooled' cosmid library. Notably, two cosmids, TM4cos7 and TM4cos20, were not able to recombine with any other cosmids, even though the complete genome was represented in all pairs tested (Table 13). However, it is interesting that small numbers of plaques were obtained with these pairs, whereas zero plaques were consistently obtained with pairs that were not expected to recombine. This deficiency in recombination is likely due to the presence of a *cis*-acting element in this region that is required for recombination and/or DNA replication. The region encodes only one small gene in entirety, *71*, but does also include the 3′ half of gene *70*, which is predicted to encode the DNA primase. Therefore, the *cis*-acting element could be an origin of replication, which is often located in the region of the genome that encodes DNA replication proteins in other phages and in bacteria [58,99].

Table 13. Recombination between TM4 cosmids as measured by plaque formation.

Individual		Pairs w/ overlapping deletions		Pairs w/ non-overlapping deletions		TM4Cos7 and TM4Cos20 pairs	
TM4Cos#[a]	Pfu per μg DNA[b]	TM4Cos#[a]	Pfu per μg DNA[b]	TM4Cos#[a]	Pfu per μg DNA[b]	TM4Cos#[a]	Pfu per μg DNA[b]
7	0	8 + 11	0	8 + 9	235	7 + 8	2
8	0	9 + 49	0	9 + 11	353	7 + 9	3
9	0	9 + 53	0	9 + 12	252	7 + 11	0
11	0	14 + 13	0	9 + 42	263	7 + 14	2
12	0			8 + 14	491	14 + 20	2
13	0			11 + 49	248		
14	0						
20	0						
42	0						
49	0						
53	0						
TM4 DNA (pfu per μg)[c]	1×10^4						
Pooled DNA (pfu per μg)	320						

a. Each cosmid was assigned a number during screening: TM4cosX.
b. The number of plaques (pfu) per μg total DNA is shown from transformations with DNA from either single cosmids, pairs of cosmids, or a pooled library.
c. Wild type TM4 DNA (200 ng) was used as a positive control and is represented as pfu/μg.

Analysis of plaques resulting from transformations with pooled cosmid DNA or pairs of cosmids showed only the presence of wild type TM4 DNA (Figure 33A). These recombinant plaques did not show the presence of the *E. coli* plasmid (assayed by PCR), and DNA prepared from the plaques displayed a restriction pattern identical to wild type (Figure 33B). No true shuttle phasmids were identified out of 14 plaques screened from the pooled library, which is expected since these were recovered from the previous study at a very low frequency (~0.25%). Further, the average size of the deletions in the cosmids was ~9 kbp, which is much larger than the deletions found in shuttle phasmids, such as phAE87 (305 bp) and phAE159 (~5856 bp) [14], and is therefore more likely to have removed essential genes.

Recombination between the TM4 cosmids could conceivably be derived from either host or phage recombination protein activity. Therefore, similar assays were performed in Δ*recA* and Δ*recB M. smegmatis* strains (gifts of K.G. Papavinasasundaram and K. Derbyshire, respectively) to determine the role of host recombination. Recombinant wild type plaques were obtained in both strains using pairs of cosmids to assess recombination (Figure 33C), but unpredictably, recombination levels were consistently higher in the Δ*recA* and Δ*recB* strains (~2-fold) compared to wild type. These data demonstrate that TM4 cosmid molecules recombine *in vivo* to yield wild type TM4 DNA independently of host RecA and RecB, which suggests that TM4 encodes a recombination system.

Figure 33. TM4 cosmids recombine *in vivo* to yield wild type TM4, independently of host RecA and RecB.

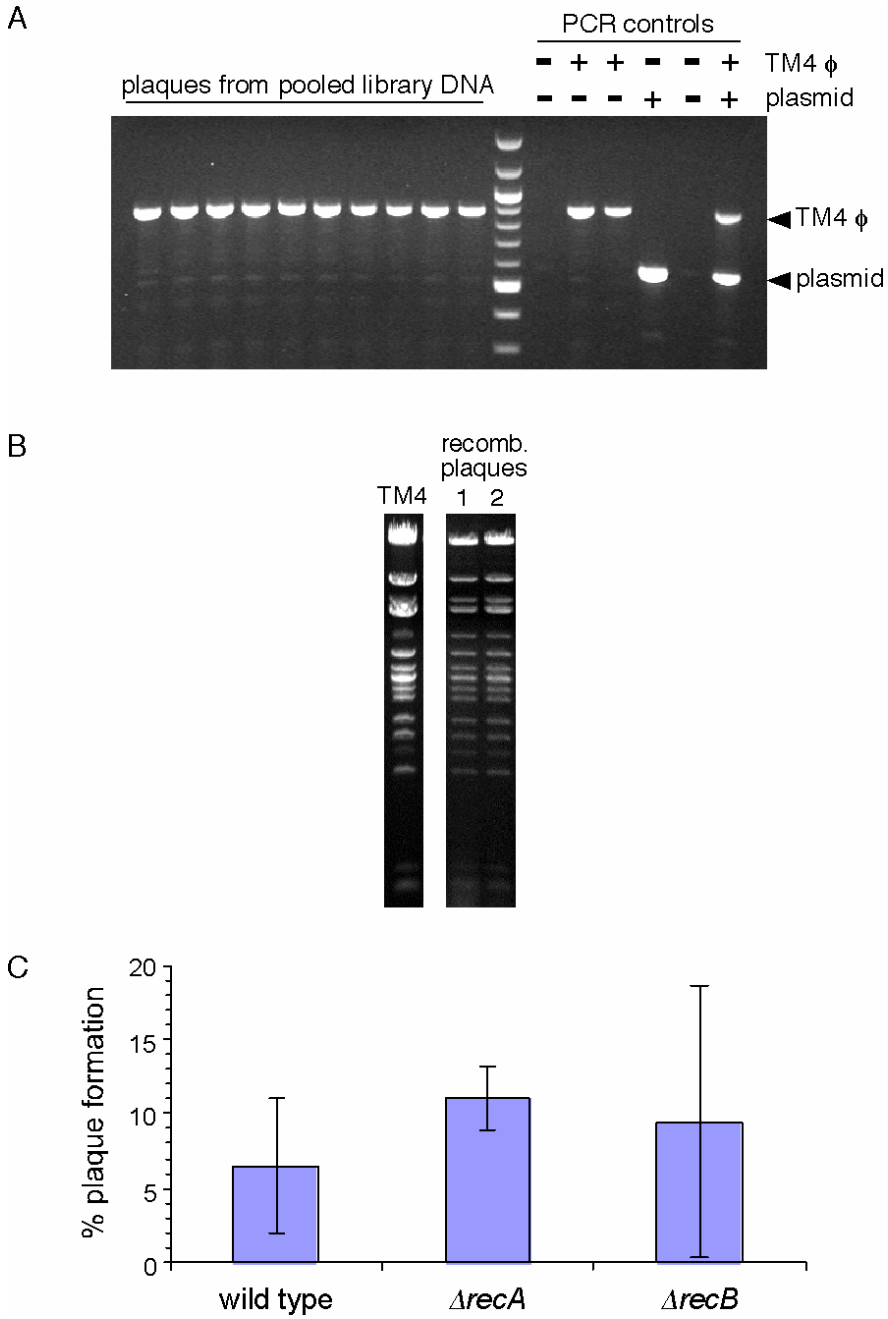


Figure 33. (A) Plaques from transformations with DNA from the pooled cosmid library were analyzed by PCR with two sets of primers that amplify TM4 DNA (880 bp) and pYUB854 DNA (584 bp). Controls included (from left to right): a plug from a lawn of *M. smegmatis*, a TM4 plaque, TM4 DNA, pYUB854 DNA, no DNA, and TM4cos11 DNA. (B) DNA was prepared from two recombinant plaques and analyzed by *BstE*II restriction digest alongside wild type TM4 DNA as a control. (C) Cosmid pairs with non-overlapping deletions (*e.g.* TM4cos9 and TM4cos11) were co-transformed into wild type, Δ*recA*, and Δ*recB M. smegmatis* strains, and plaque numbers were recorded; TM4 DNA was transformed separately as a positive control. For each transformation, the number of plaques from cosmid transformations (per μg) was divided by the number of TM4 plaques (per μg) and represented as percent plaque formation. The data shown represent the average of eight independent experiments, with error bars calculated from standard deviations.

A number of other experimental approaches for identifying the TM4 recombination proteins were attempted without success. First, an *M. smegmatis* TM4 genomic library was constructed in which TM4 fragments were cloned in the pLAM12 vector under control of $P_{acetamidase,}$ and a library of these were transformed into wild type *M. smegmatis*. These cells were induced for expression and prepared similarly to Che9c gp60/gp61-expressing cells. Recombination activity was assayed by ssDNA recombineering using oligonucleotides that introduce point mutations that confer drug-resistance. However, transformation of the pooled library cells did not produce recombinant colonies in duplicate experiments. Therefore, as a second approach, individual segments of the TM4 genome (~3 kbp, excluding known structural genes) were cloned in pLAM12. However, only two out of ten plasmids successfully transformed *M. smegmatis*. This suggests that even the leaky expression of the acetamidase promoter is sufficient to cause toxicity with some of these genes, and therefore a different promoter or vector may be required. A similar result that may be a result of leaky expression was observed in experiments with Halo genes *41-44* cloned under the acetamidase promoter. It was observed that constructs that expressed Halo gp41-44 on a replicating vector (pLAM12 parent) grew very slowly, while an integrated version was better tolerated (data not shown). These experiments – both the *M. smegmatis* library and the individual TM4 clones – could be repeated in a different vector background.

4.5 CONCLUSIONS

4.5.1 Mycobacteriophage-encoded recombination systems

Thus far, only seven known or putative recombination systems have been identified by bioinformatic analyses in mycobacteriophages and prophages out of 51 sequenced phages and all mycobacterial sequences in the NCBI database. However, more in-depth PSI-BLAST analysis using putative recombination proteins from phages of related bacteria as queries may uncover additional mycobacteriophage genes. The observation that TM4 likely encodes a recombination system that is not recognizable by sequence similarity suggests that these proteins are probably present in other phages but are, thus far, unidentified. The putative Giles recombination system could have easily been overlooked without careful scrutiny if not for the similarity between Giles gp53 and Halo gp43, and these genes were only recognized because the gene adjacent to Halo gp43 is a recognizable RecE homologue. Approximately 50% of the mycobacteriophage ORFs do not have sequence identity to proteins with known function from other organisms, but many are similar to other mycobacteriophage-encoded genes. Therefore, identification of additional phage-encoded recombination proteins – either by bioinformatic or experimental analyses – may reveal the presence of these in more mycobacteriophages.

Among the mycobacteriophage-encoded SSAPs that were examined *in vivo*, Che9c gp61 demonstrated the highest level of recombineering activity, and Halo gp43 functioned less well with an 8- to 30-fold reduction in activity. Surprisingly, however, Giles gp53 did not produce recombinant colonies above background levels, even though it shares 30% amino acid identity with Halo gp43. However, since protein expression was not confirmed, these results are not conclusive. It would therefore be interesting to examine the other putative SSAP proteins for

activity in mycobacteria. These results are reminiscent of those from the study by Datta *et al.* in which λ Beta and *E. coli* RecT demonstrated a stark difference in recombineering efficiency [43], even though they both are encoded by *E. coli* phages. This further supports the notion that development of genetic tools such as this may require characterization of multiple bacteriophages to increase the available phage gene pool in which to search for recombination proteins.

The role of these proteins in the mycobacteriophages is unknown, and the question of whether these proteins are essential in phages Che9c, Halo, and Giles is currently being tested in the Hatfull lab. Further, their role (if any) in phage propagation cannot necessarily be inferred based on data from other phages. The activities of these proteins vary in other phages, although it is common that recombination deficient phages are decreased in burst size [53,218]. For example, one function of the λ Red system is likely to increase DNA synthesis by generating additional circular genomes from linear concatemers [53,106], whereas the P22 system circularizes the genome upon entry into the host cell [237,238]. Further investigation of the prevalence of SSAP genes in mycobacteriophage genomes and their function *in vivo* will yield better insights into their biological relevance and diversity.

4.5.2 SSAP species-specificity

From this study and that performed by Datta and colleagues [43], it is apparent that there is a distinct difference in recombination activity when the same SSAPs are tested in *M. smegmatis* and *E. coli*. Although this could be due to experimental variation, it is more likely a result of the inherent species-specific nature of these proteins. Recombination proteins encoded by Gram-negative bacteria tended to have the highest activities in *E. coli*, whereas proteins from phages

infecting more distantly related hosts displayed decreased activity. In *M. smegmatis*, expression of Che9c gp61 facilitated the highest recombineering frequencies, with Halo gp43 and *E. coli* RecT moderately lower in activity (30- and 10-fold, respectively). Therefore, the finding that Che9c gp61 functions at a low level in *E. coli* is clearly not due to an inherently poor activity of this protein, which was suggested by Datta *et al.* [43], but instead is probably due to expression in a distantly-related organism. However, the data from the two studies correlated in the general observation that the SSAPs – specifically λ Beta and Che9c gp61 – displayed the highest activities above others tested in the native bacterial hosts of the phages from which they were derived. It is interesting to note that *E. coli* RecT had a high level of activity in *M. smegmatis*, whereas λ Beta was not active. This is in contrast to observations made in *E. coli* where λ Beta functioned substantially better than RecT in *E. coli*.

Overall, it appears that SSAP proteins function optimally in bacteria that are closely related to the hosts of their respective phages. This could be due to specific interactions with host proteins that occur during recombination. One plausible hypothesis is that the SSAP – and possibly the exonuclease – interacts with components of the DNA replication machinery, as replication is a process that has a direct effect on λ Red-, RecET-, and Che9c gp60/gp61-mediated recombination efficiency [52,113,228,244]. A potential candidate for this interacting partner is the host SSB because it is associated with ssDNA during DNA synthesis, and this is therefore the suggested target for SSAP-ssDNA complex recombination. Therefore, extension of the recombineering technology to other organisms may require identification of a recombination system encoded by a host-specific phage in order to produce functional protein interactions for optimal activity.

The SSAPs examined in this study have yet to be tested in other mycobacterial species, such as *M. tuberculosis*. Recombineering with ssDNA – though not with dsDNA – mediated by Che9c gp61 in *M. tuberculosis* is decreased 5- to 30-fold as compared to *M. smegmatis* in the same assays. This could also be due to slight differences between mycobacterial species in host protein-SSAP interactions that result in a decreased recombination efficiency. It is therefore possible that a different mycobacteriophage-encoded SSAP, such as Halo gp43, may improve recombineering frequencies in *M. tuberculosis*.

4.5.3 The TM4 recombination system

The *in vivo* recombination assay with TM4 cosmids observed by Jacobs *et al.* was repeated in this study, both with pools of the entire library of molecules and with pairs of cosmids [86]. No true shuttle phasmids were identified, though this is not surprising due to the average size of the deletions. Cosmid recombination was independent of the activities of the RecBCD complex, as well as the major host recombination protein, RecA. Therefore it appears that the activity is derived from phage-encoded proteins, although there may be other host proteins that are required. It is striking that an analysis of the TM4 genome does not reveal any pairs of proteins with sequence similarity to known recombination proteins. This suggests that the proteins required for recombination of the TM4 cosmids may be a new family of recombinases and/or exonucleases, and potentially these genes are located in separate regions of the genome. Experiments designed to screen for the recombination proteins were not initially successful, but the data provided a basis for altering the experimental setup. Further, a test screen should be performed with a phage genome – such as Che9c – that is known to encode recombination proteins. If the region of the Che9c genome that encodes gp61 can be identified in this type of

screen, this lends support to the utility of this experiment for other phages without recognizable recombination homologues.

During the course of experiments with TM4 cosmids, a putative *cis*-acting element was discovered in the region of the genome between 42,796 bp and 44,854 bp. This is most likely the location of the origin of replication. Two pieces of evidence support this hypothesis; first, this region includes the putative DNA primase gene (*70*), and in other phages – such as λ – the origin of replication is present in the region encoding genes required for DNA replication [58]. Second, DNA replication plays a critical role in recombination in the single strand annealing pathway – such as with λ Red – by providing recombinogenic substrates [215,216,222]. Phage proteins required for DNA replication should be provided in *trans* by the other cosmid, but the deletion of the cosmid origin of replication cannot be compensated and may severely limit recombination. It is also possible that recombination still occurs at low levels, as evidenced by the small numbers of plaques observed in recombination assays with cosmids deleted for this region. Further experimentation would be required to clearly identify the TM4 origin of replication and/or this *cis*-acting element.

5.0 DISCUSSION

5.1 MYCOBACTERIAL RECOMBINEERING

Bacteriophages have long demonstrated their utility for advancing tools for genetics and molecular biology in their bacterial hosts. Some of the more well-known examples of this are DNA ligase, T4 polymerase, and various restriction enzymes. This is further exemplified by the use of phage-encoded recombination proteins for the recombineering technology originally developed in *E. coli*, and later extended for use in other Gram-negative bacteria. The mycobacteriophages are no exception; the sequencing and characterization of these phages has provided a vast reservoir of genes to study and exploit for materials such as integration-proficient plasmids, selectable markers, and most recently, the mycobacterial recombineering system. The development of this system will allow members of the mycobacterial research community to perform genetic manipulations with an efficiency that is unparalleled by any other technique. Gene replacement mutagenesis by recombineering requires the same amount of DNA cloning and cell preparation as the minimum amount required for any other technique. Construction of the AES merely requires the standard synthesis of a linear substrate with ~500 bp homology flanking an antibiotic resistance cassette. No further manipulations of the AES are required, nor are the rounds of screening needed for some methods, since 90% of the mutants generated by recombineering are correctly targeted. Electrocompetent cell aliquots of the mycobacterial

recombineering strain (containing plasmid pJV53 or a similar construct) can be prepared in advance and stored, which minimizes experimental preparation. Further, recombineering of point mutations does not require any plasmid construction, since the short ssDNA substrates can be synthesized commercially. Importantly, mutations that are not directly selectable can be made at a relatively high frequency (3-5%) by using a co-transformation technique. Removal of the recombineering plasmid can also be simplified by using a *sacB* gene for counter-selection. Another potential use of recombineering is the deletion of sequences, such as entire genes, internal domains, or the antibiotic resistance genes in marked gene replacement mutants. This has been demonstrated by deleting most of the *M. smegmatis leuD* gene, and likely can be used for other purposes.

5.1.1 Future applications of mycobacterial recombineering

The mycobacteriophage Che9c-encoded recombination system has provided a means for improving genetic techniques in mycobacteria, and it is likely that further extension of this technology will be made for other purposes. Targeted gene replacement mutagenesis has obvious potential for making complete gene deletion sets for *M. tuberculosis* and *M. smegmatis*, a feat that otherwise would be too time-consuming with available methods. Not only would this provide mutant strains for various experimental purposes, but it would also supplement the data pertaining to gene essentiality from previous genome-wide studies [200].

Additionally, nonsense mutations could be introduced into putative essential genes to assay essentiality and gene function. The initial experimental design to test this approach involved the use of a nonsense codon suppressor tRNA gene derived from mycobacteriophage L5, which has been shown previously to suppress amber mutations in mycobacteria [60]. The

amber suppressor gene has been cloned such that its expression should be controlled by the Tet inducible promoter [51], although suppression of amber mutations has thus far been unsuccessful. However, alternative expression systems could be tested that would provide tightly controlled induction or repression. Nonsense mutations could then be introduced into test genes by ssDNA recombineering, and the viability of the mutant strain could be assessed in the presence or absence of nonsense suppressor gene expression. Although this approach has yet to be tested, it offers the potential for analysis of gene essentiality at an individual locus or genome-wide level. Finally, mutagenesis by ssDNA recombineering allows point mutations to be inserted in isogenic strains for direct and uncomplicated comparisons. This is more beneficial compared to previous methodologies that typically required gene deletion followed by complementation, and therefore analyses were not performed under endogenous conditions. This can be specifically applied for determining the role of mutations that confer drug-resistance, which may aid in research on the origins of XDR *M. tuberculosis* strains.

The extension of this technology for mutagenesis of mycobacteriophage genomes recently has provided a simple method for future genomic and proteomic study of phages (L. Marinelli, manuscript in preparation). For example, current experiments are testing if recombineering can be used to insert His-tags onto phage genes to facilitate simple purification of tagged proteins directly from infected cells. Several phages containing either point mutations or deletions have been constructed and are also currently being studied. In addition, gene essentiality can be tested, which has been demonstrated in a proof-of-principle experiment involving the deletion of the *lysA* gene of mycobacteriophage Giles.

5.2 MYCOBACTERIOPHAGE-ENCODED RECOMBINATION PROTEINS: A MODEL FOR DEVELOPMENT OF A RECOMBINEERING SYSTEM

The mycobacteriophages are a fascinating group of organisms that have greatly contributed to our knowledge of evolution, morphologic and genetic diversity, biochemisty, and the biological consequences of phage-host interactions. Phage genome sequencing contributes to the expanding gene pool, a useful source for studies of gene function and the development of genetic tools. At the beginning of this project, the only sequenced mycobacteriophages encoding putative homologous recombination systems were Che9c and Halo, and therefore only these were available for study. Subsequently, phages BPs and Giles were sequenced, and similar proteins were identified, while careful PSI-BLAST analyses continue to reveal additional putative recombinases. Also of interest are the prophages that appear to be present in the genomes of *M. avium* and *M. abscessus* and encode SSAP recombinase homologues, as well as mycobacteriophages Wildcat and Cjw1. Although none of these proteins were examined any further in this study, they are also potential candidates for recombinase activity *in vivo*.

Interestingly, the mycobacteriophage-encoded recombinases that were tested had varying levels of activity in the *M. smegmatis* ssDNA recombineering assay. Fortuitously, the first recombinase used, Che9c gp61, exhibits the highest levels of recombination activity *in vivo* thus far. Halo gp43 is slightly less efficient, and Giles gp53 did not show any activity in these assays, although in this case expression was not confirmed. This first suggests that identification of only one phage-encoded recombination protein may not be sufficient for development of a recombineering system in other bacteria. Instead, these findings support the notion that identification and analysis of multiple phage-encoded proteins is preferable in order to optimize recombineering frequencies. In light of the species-specificity observed for both the *E. coli* and

mycobacterial phage-encoded recombination proteins, it is clear that optimal levels of recombineering can best be achieved through isolation and sequencing of host-specific bacteriophages. Therefore, it is likely that recombineering systems can be developed in virtually any genetically tractable bacterium for which at least basic genetic tools – such as plasmids and expression cassettes – have been described.

An important consideration is the question as to why the mycobacteriophage-encoded recombinases display varying levels of activity in *M. smegmatis*, particularly since they all appear to belong to the same SSAP superfamily. One attractive explanation is that they each function optimally in the preferred host bacterium of their respective phages. The species-specific nature of these proteins – observed broadly between phage-encoded proteins of distantly-related host bacteria such as *E. coli* and *M. smegmatis* – likely affects activity even in closely related bacteria of the same genus. Che9c gp61, for example, shows decreased recombination efficiency in *M. tuberculosis* compared to *M. smegmatis*. The basis of the differing activity levels may be attributed to specific recombinase-host protein interactions – during processes such as DNA replication – that are required for optimal recombineering. The role of replication is particularly interesting to consider with regard to the fast- and slow-growing mycobacteria. Although this is not well-studied, the rate, processivity, and/or regulation of DNA replication in *M. tuberculosis* is probably dissimilar to *M. smegmatis*, which may have a profound effect on recombineering frequencies. Therefore, perhaps expression of another recombinase will be more suitable for recombineering in *M. tuberculosis* or other mycobacteria. Halo gp43 is a particularly interesting candidate because Halo can infect *M. tuberculosis* (T. Sampson, personal communication). This logic can also be applied to other bacteria. While the λ

Red proteins may function sufficiently in some Gram-negative bacteria, developing recombineering in others may require testing of additional host-specific phage-encoded systems.

It is likely that additional homologues of known recombination proteins will be found as more phages and bacteria are sequenced. Possibly more interesting, however, are the phage genes that are not detectably related to known recombinases, but still function similarly. The genes that encode the recombination system of TM4 remain anonymous, and even a recent analysis did not reveal any likely candidates. Clearly, a screen will be necessary to identify these proteins. This tactic could then be used to develop recombineering in any bacterial system for which phages have been isolated but recombination proteins are not recognizable (or if the phage is not sequenced). The simplest approach appears to be the construction of a phage genomic library in several different vector backbones (integrating or replicating), potentially with different promoters in order to test varying expression levels. Subsequently, the library of bacterial cells containing these plasmids would be screened for activity using ssDNA recombineering of an allele conferring a drug-resistant phenotype. To test this, a screen should be performed first using a phage genome that is known to encode recombination proteins, such as Che9c or Halo. If this is successful, it would lend support to the use of this approach as a broadly applicable method for identifying phage recombinases, potentially one that could be used for phages of other bacteria.

6.0 MATERIALS AND METHODS

6.1 REAGENTS AND BUFFERS

6.1.1 Growth media

7H9 broth: 4.7 g Middlebrook 7H9 powder (Difco) was dissolved in 900 ml dH_2O and 5 ml 40% glycerol. This was autoclaved, and 100 ml ADC (see below), 2.5 ml 20% Tween 80 (if desired), and antibiotics were aseptically added as required. For growth of *M. tuberculosis*, 5 ml oleic acid per liter was added.

7H9 induction medium: 4.7 g 7H9 powder (Difco) was dissolved in 900 ml dH_2O and 5 ml 40% glycerol. This was autoclaved, and 100 ml dH_2O, 10 ml 20% succinate, 2.5 ml 20% Tween, and Kanamycin (see below) were aseptically added.

7H10 agar: 19 g Middlebrook 7H10 powder (Difco) was dissolved in 900 ml dH_2O, and 12.5 ml 40% glycerol and 4 drops anti-bubble (Pourite) were added. This was autoclaved, and 100 ml ADC and antibiotics as required were aseptically added. For growth of *M. tuberculosis*, 5 ml oleic acid per liter was added.

7H11 agar: 21 g Middlebrook 7H11 powder (Difco) was dissolved in 900 ml dH_2O, and 12.5 ml 40% glycerol and 4 drops anti-bubble (Pourite) were added. This was autoclaved, and 100 ml

ADC, plus 5 ml oleic acid (or 100 ml OADC, BDL), and antibiotics were aseptically added as required.

Mycobacterial top agar (MBTA): 4.7 g Middlebrook 7H9 powder (Difco) and 7 g Bacto Agar were dissolved in 900 ml dH_2O and autoclaved.

ADC: 20 g dextrose and 8.5 g NaCl were dissolved in 950 ml dH_2O. 50 g Albumin (Spectrum Biochem) was added and stirred with no heat until dissolved. This was filter-sterilized through a 0.22-μm-pore membrane and stored at 4°C.

20% Tween 80: Tween 80 was dissolved at 20% (v/v) by heating to 56°C, filtered through a 0.22-μm-pore membrane, and stored at 4°C. This was used at a final concentration of 0.05% in liquid media.

20% acetamide: Acetamide (Sigma) was dissolved at 20% in dH_2O, filtered through a 0.22-μm-pore membrane, and stored at 4°C. This was used at a final concentration of 0.2% in media.

20% succinate: Sodium succinate dibasic hexahydrate (succinic acid, Sigma S9637) was dissolved at 20% in dH_2O, filtered through a 0.22-μm-pore membrane, and stored at 4°C. This was used at a final concentration of 0.2% in media.

Oleic acid: Oleic acid (Sigma) was dissolved at 10 mg/ml in dH_2O by heating the ampule in 37°C water bath and stirring into dH_2O with heat until completely dissolved. 1 g NaOH was added and stirred until dissolved, and the solution was filtered through a 0.22-μm-pore membrane and stored in 10 ml aliquots at -20°C. Used at a final concentration of 50 μg/ml in *M. tuberculosis* media.

Luria-Bertani broth (LB broth): 20 g LB broth (Difco) was dissolved in 1 L dH_2O. This was autoclaved and antibiotics were added aseptically when required.

Luria-Bertani agar (LB agar): 35 g LB agar (Difco) was dissolved in 1 L dH_2O, and 4 drops anti-bubble (Pourite) were added. This was autoclaved, and antibiotics were added aseptically when required.

Tryptic Soy Broth (TSB): 30 g TSB (Difco) was dissolved in 1 L dH_2O. This was autoclaved, and antibiotics were added aseptically when required.

6.1.2 Antibiotics and Supplements

Carbenicillin: (Cb, Sigma) was dissolved at 50 mg/ml in dH_2O, filtered through a 0.22-μm-pore membrane, and stored at 4°C.

Chloramphenicol: (CM, Sigma) was dissolved at 100 mg/ml in 100% ethanol and stored at 4°C.

Cycloheximide: (Chx, Sigma) was dissolved at 10 mg/ml in dH_2O, filtered through a 0.22-μm-pore membrane, and stored at 4°C.

Ethambutol: (EMB, Sigma) was dissolved at 50 mg/ml in dH_2O, filtered through a 0.22-μm-pore membrane, and stored at 4°C.

Ethionamide: (ETH, Sigma) was dissolved at 50 mg/ml in 100% DMSO and stored at 4°C.

Gentamicin: (Gent, sulfate salt, Sigma) was dissolved 10 mg/ml in dH_2O, filtered through a 0.22-μm-pore membrane, and stored at -20°C in 1 ml aliquots.

Hygromycin B: (Hyg, Sigma) was dissolved at 100 mg/ml in dH_2O, filtered through a 0.22-μm-pore membrane, and stored at -20°C in 1 ml aliquots.

Isoniazid: (INH, isonicotinic hydrazide, Sigma) was dissolved at 50 mg/ml in dH_2O, filtered through a 0.22-μm-pore membrane, and stored at 4°C. Solutions were made fresh and used within one week.

Isopropyl β-D-1-thiogalactopyranoside: (IPTG) was resuspended to either 1 M or 0.1 M in dH_2O, filtered through a 0.22-μm-pore membrane, and stored at 4°C.

Kanamycin: (Kan, Sigma) was dissolved at 50 mg/ml in dH_2O, filtered through a 0.22-μm-pore membrane, and stored at 4°C.

Leucine: (LEU, Sigma) was dissolved at 10 mg/ml in dH_2O, filtered through a 0.22-μm-pore membrane, and stored at 4°C.

Ofloxacin: (OFX, Sigma) was dissolved at 50 mg/ml in 1 N NaOH, filtered through a 0.22-μm-pore membrane, and stored at 4°C. Solutions were made fresh and used within one week.

Pantothenate: DL-Pantothenic acid (PAN, Sigma) was dissolved at 100 mg/ml in dH_2O, filtered through a 0.22-μm-pore membrane, and stored at -20°C.

Rifampicin: (RIF, Sigma) was dissolved at 50 mg/ml in 100% DMSO and stored at 4°C wrapped in foil. Solutions were made fresh and used within one week.

Streptomycin: (Str, Sigma) was dissolved at 50 mg/ml in dH_2O, filtered through a 0.22-μm-pore membrane, and stored at 4°C. Solutions were made fresh and used within one week.

Tetracycline: (Tet, Sigma) was dissolved at 5 mg/ml in dH_2O, filtered through a 0.22-μm-pore membrane, and stored at -20°C. Tetracycline is light sensitive – media was prepared only at the time needed.

Uracil: (URA, Sigma) was dissolved at 200 mM (22.41 mg/ml; 112.1 MW) in 1 N NaOH, filtered through a 0.22-μm-pore membrane, and stored at 4°C.

X-gal: 100 mg (entire bottle from Invitrogen) was dissolved at 50 mg/ml in 100% DMF (2 ml), wrapped in foil and stored at -20°C.

5-Fluoro-orotic acid: (5-FOA, US Biologicals) was added to agar media prior to autoclaving at 1 mg/ml. Powder goes into solution throughout autoclave cycle.

6.1.3 Laboratory reagents and stock solutions

0.1 M $CaCl_2$: 11.1 g $CaCl_2$ was dissolved in 1 L dH_2O and autoclaved.

100X Denhardt's solution: 2 g BSA, 2 g Ficoll, and 2 g Polyvinyl pyrrolidine were dissolved in 100 ml dH_2O, filtered through a 0.22-μm-pore membrane, and stored at -20°C.

0.1 M DTT: Dithiothreitol was dissolved in dH_2O at 0.1 M (1.54 g per 10 ml; 154 MW), filtered through a 0.22-μm-pore membrane, and stored at -20°C in 1 ml aliquots.

0.5 M EDTA: 93.06 g disodium EDTA (Na_2EDTA) was dissolved in 400 ml dH_2O on low heat, and pH was adjusted to 7.5 with NaOH (~50 ml 0.5M NaOH). Volume was brought to 500 ml and autoclaved.

0.1 M Na_2EDTA pH 8.8: 37.2 g of disodium EDTA (Fisher) was dissolved in 1 L dH_2O, and pH was adjusted to 8.8 with NaOH.

40% Glycerol: 400 ml Glycerol was mixed in 600 ml dH_2O and autoclaved.

40% Glucose: 400 g dextrose was dissolved in 750 ml dH_2O and autoclaved. Water was added first and dextrose was slowly added to dissolve.

0.25 M HCl: Stock solution HCl (11 M) was diluted to 0.25 M by adding 22.7 stock HCl per one L of dH_2O.

0.5 N KOH: 28.1 g of KOH was dissolved in 1 L dH_2O and stored at room temperature.

1 M $MgCl_2$: 203 g $MgCl_2$ hexahydrate (or 95.21 g anhydrous) was dissolved in 1 L dH_2O and autoclaved.

1 M $MgSO_4$: 120.37 g $MgSO_4$ was dissolved in 1 L dH_2O and autoclaved.

5 M NaCl: 292 g NaCl was dissolved in 1 L dH_2O, heated on low to dissolve, and autoclaved.

3 M NaOAc: 408.1 g NaOAc was dissolved in 400 ml dH_2O on low heat and pH adjusted to 5.2 with glacial acetic acid (~200 ml). Volume was brought to 1 L and autoclaved.

5 N NaOH: 200 g of NaOH was dissolved in 1 L dH_2O.

0.4 N NaOH: 80 ml of 5 N NaOH was brought to 1 L with dH_2O.

Phage Buffer: 4 g NaCl was dissolved in 980 ml dH_2O, and 10 ml of 1 M Tris pH 7.5 and 10 mL of 1 M $MgSO_4$ were added and autoclaved.

Phenylmethanesulphonylfluoride: (PMSF) was dissolved in isopropanol to 100 mM. The tube was wrapped in foil and stored at -20°C.

Proteinase K: proteinase K was dissolved at 10 mg/ml in dH_2O, filtered through a 0.22-μm-pore membrane, and stored at -20°C in 1 ml aliquots.

10% SDS: 100 g SDS was dissolved in 1 L dH_2O, filtered through a 0.22-μm-pore membrane, and stored at room temperature.

20X SSC: 175.32 g NaCl and 88.23 g sodium citrate were dissolved in 1 L dH_2O and pH was adjusted to 7.0 with HCl.

1 M Tris pH 7.5: 121 g Trizma base was dissolved in 800 ml dH_2O, and pH was adjusted to 7.5 with HCl. Volume was brought to 1 L and autoclaved.

1 M Tris pH 8.0: 121 g Trizma base was dissolved in 800 ml dH_2O, and pH was adjusted to 8.0 with HCl. Volume was brought to 1 L and autoclaved.

Tris-buffered saline ± Tween: (TBS, TBS-T) 25 mM Tris pH 8.0, 125 mM NaCl, ± 0.1% Tween.

Tris-EDTA (TE): 10 ml of 1 M Tris, pH 7.5 and 2.5 ml of 0.5 M EDTA were mixed in 987.5 ml dH_2O and autoclaved. Final solution: 10 mM Tris pH 7.5, 1.25 mM EDTA.

6.1.4 Gel electrophoresis

6.1.4.1 Agarose gel electrophoresis

20% Ficoll dye: 2 g Ficoll was heated on low heat to dissolve in 8 ml dH_2O. 1 ml 1% (w/v) bromophenol blue and 1 ml 1% (w/v) xylene cyanol were added.

Tris Borate EDTA (TBE): 121.1 g Trizma base, 51.25 g Boric Acid, and 3.72 g EDTA were dissolved in ~890 ml to make a 10X solution. 1X = 100 mM Tris (121.1 MW), 83 mM Boric Acid (61.83 MW), and 1 mM EDTA (372.24 MW).

6.1.4.2 Polyacrylamide gel electrophoresis

Coomassie Blue Stain: 2.5 g Coomassie Blue was dissolved in 450 ml methanol, and 90 ml acetic acid was added and volume brought to 1 L dH_2O. This was filtered and stored at room temperature.

Glycerol loading dye: 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, and 30% glycerol

Protein gel running buffer (10X): 144 g glycine, 30 g Trizma base, and 10 g SDS were dissolved in 1 L dH_2O by mixing SDS first, and pH was adjusted to 8.3.

4X SDS-PAGE loading dye: 125 mM Tris pH 7.5, 20% glycerol, 2% SDS, 5% β-mercaptoethanol, 0.1% bromophenol blue dye. In 8 ml, add 2.6 ml dH_2O, 1 ml 0.5 M Tris pH 7.5, 1.6 ml neat glycerol, 1.6 ml 10% SDS, 0.4 ml β-mercaptoethanol, 0.8 ml 1% bromophenol blue.

SDS-PAGE Laemmli gels: For 2 small 10% separating gels: 5 ml dH_2O, 2.5 ml 4X separating buffer, 2.5 ml 40% acrylamide:bisacrylamide (29:1), 10 µl TEMED, 40 µl 10% (w/v) APS. For 2 small stacking 4.5% stacking gels: 3 ml dH_2O, 1.25 ml 4X stacking buffer, 0.5 ml 40% acrylamide:bisacrylamide (29:1), 15 µl TEMED, 25 µl 10% APS.

4X SDS-PAGE separating buffer: 18.17 g Trizma base was dissolved in ~80 ml dH_2O, and 4 ml 10% SDS was added. pH was adjusted to 8.8 with HCl (~2.5 ml), and volume was brought to 100 ml.

4X SDS-PAGE stacking buffer: 6.06 g Trizma base was dissolved in ~80 ml dH_2O, and 4 ml 10% SDS was added. pH was adjusted to 6.8 with HCl, and volume was brought to 100 ml.

TBE polyacrylamide native gels: For 1 large 8% gel: 42 ml dH_2O, 6 ml 10X TBE buffer, 12 ml 40% acrylamide:bisacrylamide (29:1), 21 µl TEMED, 420 µl 10% APS.

6.1.5 Assay buffers

Annealing buffer: 50 mM Tris pH 7.5, 100 mM NaCl

Binding assay reaction buffer: 33 mM Tris pH 7.5, 13 mM $MgCl_2$, 1.8 mM DTT, 88 µg/ml BSA

Binding assay wash buffer: 33 mM Tris pH 7.5, 13 mM $MgCl_2$, 1.8 mM DTT

Exonuclease assay buffer: 20 mM Tris, pH 8.0, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol

Genomic DNA prep CTAB solution: 4.1 g NaCl was dissolved in 90 ml dH_2O, and 10 g Cetrimide was added, stirring. This was incubated at 65°C until in solution, and stored at room temperature.

Gel filtration equilibration buffer: 33 mM Tris pH 7.5, 100 mM NaCl

Genomic DNA prep GTE solution: 25 mM Tris pH 8.0, 10 mM EDTA, 50 mM glucose

Protein dilution buffer: 10 mM Tris pH 7.5, 1 mM DTT, 1 mg/ml BSA

Protein purification lysis buffer: 50 mM Tris pH 8.0, 300 mM NaCl, 5% glycerol, 1 mM PMSF

Protein purification wash buffer: 50 mM Tris pH 8.0, 300 mM NaCl, 5% glycerol, 0-20 mM imidazole

Protein purification elution buffer: 50 mM Tris pH 8.0, 300 mM NaCl, 5% glycerol, 20-200 mM imidazole

Protein storage buffer: 20 mM Tris pH 8.0, 0.1 mM EDTA, 0.1 mM DTT, 50 mM NaCl, 50% glycerol

Southern blot pre-hybridization buffer: 6X SSC, 2X Denhardt's solution, 0.1% SDS

Southern blot hybridization buffer: 6X SSC, 20 mM $NaPO_4$, pH 7.5, 5% PEG 8000 (Sigma)

Southern blot wash buffer 1: 2X SSC, 0.1% SDS

Southern blot wash buffer 2: 0.2X SSC, 0.1% SDS

Western blot transfer buffer: 48 mM Tris, 39 mM Glycine, 0.037% SDS, 20% methanol

6.2 PLASMID CLONING

6.2.1 Plasmid maintenance in *E. coli* strains

Plasmid constructs were maintained in either *E. coli* DH5α or GC5 (GeneChoice; similar genotype to DH5α) strains. When required, BL21(DE3) pLysS cells (Invitrogen) were used for

protein over-expression plasmids. Plasmids were transformed into chemically competent *E. coli* strains as described below (*E. coli* transformations). *E. coli* strains containing plasmids were mixed with sterile glycerol at a final concentration of 20% and stored at -80°C.

6.2.2 Plasmids

All of the plasmids that I made are described in Table 15. The genes of interest that were cloned, the parental insert source (PI), and the primers or restriction sites used for the insert are listed. The plasmid backbone that was cloned into (parental plasmid, PP), the restriction sites used in the vector backbone are listed, and other pertinent information is also included. For plasmids that were available commercially or were obtained from other lab members or collaborators, brief descriptions of the pertinent aspects of the plasmid (such as antibiotic resistance cassettes, genes of interest, etc.) are included in Table 14.

Table 14: Plasmids constructed by others.

Plasmid name	Features	Reference
p0004S	Cloning vector containing Hyg^R-sacB cassettes flanked by MCSs and γδ res sites, oriE, λ *cos* packaging site, oriE	Gift from W.R. Jacobs, Jr.
p0004S:leuB	*M. smegmatis leuB* KO plasmid containing upstream and downstream homology to the *leuB* locus in p0004S	Gift from W.R. Jacobs, Jr.
p0004S:leuD	*M. smegmatis leuD* KO plasmid containing upstream and downstream homology to the *leuD* locus in p0004S	Gift from W.R. Jacobs, Jr.
pAVN30	Contains P_{hsp60}-*sacB* cassette for negative selection, Hyg^R	Gift from W.R. Jacobs, Jr.
pBluescript SK+	*E. coli* cloning vector, oriE	Stratagene
pBR322	*E. coli* cloning vector, Tet^R, Amp^R, oriE	NEB
pBRL301	Contains "new" modified P_{hsp60}-*sacB* cassette, Hyg^R, oriE	Gift from W.R. Jacobs, Jr.

pET21a	T7 expression vector carrying a C-terminal 6x His-tag, Amp^R, oriE	Novagen
pET28c	T7 expression vector carrying an N-terminal 6x His-tag, Amp^R, oriE	Novagen
pGH542	Expresses γδ resolvase constitutively for unmarking gene knockouts, Tet^R, oriE, oriM	Gift from G. Hatfull
pGH1000A	Derivative of pMOSHyg containing Giles *attP-int*	[126]
phAE87	Shuttle phasmid for specialized transduction; use with pYUB854, oriE, Amp^R	[14]
pJL37	*M. bovis* BCG *hsp60* promoter, Kan^R, oriE, oriM	[18]
pJL37-P*hsp60*	Derivative of pJL37 without P*hsp60*	[61]
pKP134	*M. smegmatis pyrF* KO plasmid containing the *pyrF* gene interrupted by a $Gent^R$ cassette, oriE	[155]
pLAM12	Derivative of pJL37 containing the P*acetamidase* promoter in place of the P*hsp60* promoter	[227]
pLC3	T7 expression plasmid carrying an N-terminal 6x His-tag, MBP-fusion, and TEV protease cleavage site, Kan^R, oriE	Gift from J. Sacchetini
pLT193B-B	Contains tRNA amber suppressor cassette (in L5 gene *9*), Hyg^R, Kan^S, oriE	Gift from C. Peebles; [61]
pMOSBlue	Cloning vector containing T7-*lacZ* for blue-white screening, Amp^R, oriE	GE Healthcare
pMOSHyg	Derivative of pMOSBlue, Hyg^R	[227]
pMP6	*M. smegmatis* MSMEG0642 KO plasmid containing upstream and downstream homology to the 0642 locus in pYUB854	[172]
pMPambar	$P_{myc1tet}$-*tRNA* amber suppressor cassette, L5 *attP-Int*, $Gent^R$, oriE	Gift from M. Piuri
pMsgroEL1KO	*M. smegmatis* MSMEG4308 KO plasmid containing upstream and downstream homology to the 4308 locus in pYUB854	Gift from A. Ojha
pMsgroEL1KO	*M. smegmatis groEL1* KO plasmid containing upstream and downstream homology to the *groEL1* locus in pYUB854	[149]
pMtbgroEL1KO	*M. tuberculosis* H37Rv *groEL1* KO plasmid containing upstream and downstream homology to the *groEL1* locus in pYUB854	[227]
pMV261-lac	P_{hsp60}-*lacZ* , Kan^R, oriE, oriM	[217]
pPJM04	*M. smegmatis* MSMEG6008 KO plasmid containing upstream and downstream homology to the 6008 locus in pYUB854	Gift from P. Morris
pRDK557	*E. coli recT* gene, Kan^R	Gift from R. Kolodner
pSD26	$P_{acetamidase}$, Hyg^R, oriM	[44]
pSE100	$P_{myc1tetO}$, Hyg^R, oriE, oriM	[67]

pSJ25B	Bxb1 *attP-int*, Amp^R, oriE	[95]
pSJ25Hyg	Derivative of pSJ25B, Hyg^R	[60]
pSJ25HygSac (pPG01)	Derivative of pSJ25BHyg, sacB cassette	[60]
pTEK-4SOX	P_{smyc}-*tetR1.7*, Kan^R, oriE, oriM	[67]
pTH1-8	$P_{acetamidase}$ from pSD26, L5 *attP-Int*, Kan^R, oriE	T. Huang, unpublished data
pTTP1B	Tweety *attP-int*, Kan^R, oriE	[170]
pTX-2MIX	$P_{myc1tetO}$, P_{smyc}-*tetR*, Amp^R, oriE	Gift from S. Ehrt and D. Schnappinger
pYUB854	Hyg^R cassette flanked by MCSs and γδ res sites, oriE, λ *cos* packaging site	[14]

SDM, site-directed mutagenesis (primers listed); KO, knockout; F, forward orientation (gene cloned); R, reverse orientation (gene cloned); bl., blunted restriction site by fill-in with Klenow; Pacet, acetamidase promoter; res, resolvase; MCS, multiple cloning site; oriE, origin of replication *E. coli*; oriM, origin of replication mycobacteria.

Table 15: Plasmids constructed by JV

Plasmid Name	Gene(s) of Interest	Parental Plasmid (PP)	PP Cloning Sites Used	Parental Insert (PI) Source	PI Cloning Sites Or Primers Used	Antibiotic Markers	Mycobact. Repl/Int	Other Features
pJL37-oriM	NA	pJL37	*Mlu*I, *Xba*I, bl.	NA	NA	Kan^R	oriM	Phsp60
pMV 261-lac-amber	*lacZ* Q24*amber	pMV261-lac	NA	NA	SDM: JCV376, JCV377	Kan^R	oriM	Phsp60
pJV02F	TM4 3811-8438	pJL37	*Hpa*I	TM4	*Nae*I	Kan^R	oriM	Phsp60
pJV02R	TM4 3811-8438R	pJL37	*Hpa*I	TM4	*Nae*I	Kan^R	oriM	Phsp60
pJV03F	TM4 8438-13882F	pJL37	*Hpa*I	TM4	*Nae*I	Kan^R	oriM	Phsp60
pJV03R	TM4 8438-13882R	pJL37	*Hpa*I	TM4	*Nae*I	Kan^R	oriM	Phsp60
pJV04F	TM4 13882-17320F	pJL37	*Hpa*I	TM4	*Nae*I	Kan^R	oriM	Phsp60
pJV04R	TM4 13882-17320R	pJL37	*Hpa*I	TM4	*Nae*I	Kan^R	oriM	Phsp60
pJV05R	TM4 17320-21543R	pJL37	*Hpa*I	TM4	*Nae*I	Kan^R	oriM	Phsp60
pJV06F	TM4 21543-22750F	pJL37	*Hpa*I	TM4	*Nae*I	Kan^R	oriM	Phsp60
pJV06R	TM4	pJL37	*Hpa*I	TM4	*Nae*I	Kan^R	oriM	Phsp60

	21543-22750R							
pJV07F	TM4 22750-29001F	pJL37	*Hpa*I	TM4	*Nae*I	Kan^R	oriM	Phsp60
pJV07R	TM4 22750-29001R	pJL37	*Hpa*I	TM4	*Nae*I	Kan^R	oriM	Phsp60
pJV09F	TM4 34461-39007F	pJL37	*Hpa*I	TM4	*Nae*I	Kan^R	oriM	Phsp60
pJV09R	TM4 34461-39007R	pJL37	*Hpa*I	TM4	*Nae*I	Kan^R	oriM	Phsp60
pJV11F	TM4 42691-45427F	pJL37	*Hpa*I	TM4	*Nae*I	Kan^R	oriM	Phsp60
pJV11R	TM4 42691-45427R	pJL37	*Hpa*I	TM4	*Nae*I	Kan^R	oriM	Phsp60
pJV15	Tet^R cassette	pSJ25B	*Dra*I, *Hind*III, bl.	pBR322	*Ava*I, *Hind*III, bl.	Tet^R	Bxb1 attP/Int	
pJV16	*lacZ* 5' piece	pJV15	*BsaA*I	pMV261-lac	PCR: JCV07, JCV08	Tet^R	Bxb1 attP/Int	
pJV17	sacB cassette	pJV16	*BsaA*I	pAVN30	PCR: sacBF/R	Tet^R	Bxb1 attP/Int	sacB
pJV18	*Pml*I site	pJV17	NA	NA	SDM: JCV23, JCV24	Tet^R	Bxb1 attP/Int	
pJV19	*EcoR*V site	pTH1-8	NA	NA	SDM: JCV21, JCV22	Cb^R, Kan^R	L5 attP/Int	Pacet
pJV20	*lacZ* 3' piece	pJV18	*Pml*I	pMV261-lac	PCR: JCV09, JCV10	Tet^R	Bxb1 attP/Int	sacB
pJV21	TM4 "7-20 region"	pBR322	*BsaA*I	TM4	PCR: JCV01, JCV03	Tet^R		
pJV23	Che9c genes *59-62*	pJL37	*Hpa*I	Che9c	PCR: JCV19, JCV20	Kan^R	oriM	Phsp60
pJV24	Che9c genes *59-62*	pLAM12	*Hpa*I	Che9c	PCR: JCV19, JCV20	Kan^R	oriM	Pacet
pJV25	Che9c genes *59-62*	pJV19	*EcoR*V	Che9c	PCR: JCV19, JCV20	Kan^R	L5 attP/Int	Pacet
pJV26	Che9c genes *59-62*	pJL37-Phsp60	*Hpa*I	Che9c	PCR: JCV19, JCV20	Kan^R	oriM	
pJV27	*Msmeg recA* upstream homol.	pYUB854	*Xba*I, *Stu*I	M. smegmatis	PCR: JCV15, JCV16	Hyg^R		γδ res sites
pJV28	*Msmeg recA* downstream homol.; final recA KO plasmid	pJV27	*Hind*III, bl.	M. smegmatis	PCR: JCV17, JCV18	Hyg^R		γδ res sites
pJV29	pJV28+ shuttle phasmid phAE87	phAE87	*Pac*I	pJV28	*Pac*I	Hyg^R	shuttle phasmid	
pJV30F	TM4 "5-16kb region"	pJL37	*Hpa*I	TM4	*Kpn*I, bl.	Kan^R	oriM	Phsp60
pJV30R	TM4 "5-16kb region"	pJL37	*Hpa*I	TM4	*Kpn*I, bl.	Kan^R	oriM	Phsp60
pJV30F-	NA	pJV30F	*Hind*III,	NA	NA	Kan^R	oriM	

Hsp60			*Xba*I, bl.					
pJV30R-Hsp60	NA	pJV30R	*Hind*III, *Xba*I, bl.	NA	NA	Kan^R	oriM	
pJV31	TM4 gene *70*	pJL37	*Hpa*I	TM4	PCR: JCV13, JCV14	Kan^R	oriM	Phsp60
pJV32	TM4 gene *70*	pJL37-oriM	*Hpa*I	TM4	PCR: JCV13, JCV14	Kan^R		Phsp60
pJV33	Che9c gene *60*	pET21a	*Nde*I, *Xho*I	Che9c	PCR: JCV47, JCV48	Cb^R		C-term His tag; T7 expression
pJV34	Che9c gene *61*	pET21a	*Nde*I, *Hind*III	Che9c	PCR: JCV49, JCV50	Cb^R		C-term His tag; T7 expression
pJV35	Halo gene *42*	pET21a	*Nde*I, *Xho*I	Halo	PCR: JCV51, JCV52	Cb^R		C-term His tag; T7 expression
pJV36	Halo gene *43*	pET21a	*Nde*I, *Hind*III	Halo	PCR: JCV53, JCV59	Cb^R		C-term His tag; T7 expression
pJV37	*lacZ* 5' piece	pSJ25Hyg Sac	*Xmn*I	pMV261-lac	PCR: JCV07,JCV08	Hyg^R	Bxb1 attP/Int	sacB
pJV38	*lacZ* 3' piece	pJV37	*Sap*I, bl.	pMV261-lac	PCR: JCV09, JCV10	Hyg^R	Bxb1 attP/Int	sacB
pJV39	L5 attP/Int cassette, R	pMOSHyg	*Dra*I	pMH94	SalI, bl.	Hyg^R	L5 attP/Int	Interr. lacZ
pJV40	TM4 "5-16kb region"	pSJ25Hyg	*Kpn*I	TM4	*Kpn*I	Hyg^R	Bxb1 attP/Int	
pJV41	Halo gene *42*	pET28c	*Nde*I, *Xho*I	pJV35	*Nde*I, *Xho*I	Cb^R		N and C-term His tag; T7 expression
pJV42	Halo gene *42*	pLC3	*Nde*I, *Xho*I	pJV35	*Nde*I, *Xho*I	Cm^R		MBP-fusion
pJV43	$Gent^R$ cassette, R	pJL37	*BamH*I	pKP134	*BamH*I	Kan^R, $Gent^R$	oriM	Phsp60
pJV44	$Gent^R$ cassette, R	pJL37	*Spe*I, *Nhe*I, bl.	pKP134	*BamH*I, bl.	$Gent^R$	oriM	Phsp60
pJV45	$Gent^R$ cassette, R	pJV39	*Aat*II, *Cla*I, bl.	pKP134	*BamH*I, bl.	$Gent^R$	L5 attP/Int	Interr. Hyg^R
pJV46	Che9c genes *59-62, R*	pMOS Blue	*Eco*RV	pJV23	PCR: JCV19, JCV20	Cb^R		Interr. lacZ
pJV47	Halo genes *41-44, R*	pMOS Blue	*Eco*RV	Halo	PCR: JCV61, JCV62	Cb^R		Interr. lacZ
pJV48	sacB cassette, R	pJV24	*Spe*I	pAVN30	PCR: sacBF/R	Kan^R	oriM	Pacet, sacB
pJV49	Che9c genes *59-62*	pJL37-oriM	*Hpa*I	pJV23	PCR: JCV19, JCV20	Kan^R		Pacet
pJV50	sacB cassette, R	pJL37	*Spe*I, bl.	pAVN30	PCR: sacBF/R	Kan^R	oriM	Phsp60, sacB
pJV51	sacB cassette, R	pLAM12	*Spe*I, bl.	pAVN30	PCR: sacBF/R	Kan^R	oriM	Pacet, sacB
pJV52	Che9c gene *61*	pLAM12	*Hpa*I	pJV23	PCR: JCV49, JCV78	Kan^R	oriM	Pacet
pJV53	Che9c genes *60-61*	pLAM12	*Hpa*I	pJV23	PCR: JCV47, JCV78	Kan^R	oriM	Pacet
pJV54	Che9c gene	pJV50	*Nhe*I, bl.	pJV23	PCR: JCV49,	Kan^R	oriM	Phsp60,

	61				JCV78			sacB
pJV55	Che9c gene *60*	pLAM12	*Nde*I, *EcoR*I	pJV23	PCR: JCV47, JCV77	Kan^R	oriM	Pacet
pJV56	Che9c gene *60*/acet	pJV54	*Not*I, *Xba*I, bl.	pJV61	*Dra*I, *Nhe*I, bl.	Kan^R	oriM	Pacet-60; Phsp60-61; sacB
pJV57	Halo genes *41-44*	pJL37	*Hpa*I	Halo	PCR: JCV61, JCV62	Kan^R	oriM	Phsp60
pJV58	Halo genes *41-44*	pLAM12	*Hpa*I	Halo	PCR: JCV61, JCV62	Kan^R	oriM	Pacet
pJV59	Halo genes *41-44*	pJV19	*EcoR*V	Halo	PCR: JCV61, JCV62	Kan^R	L5 attP/Int	Pacet
pJV60	Halo genes *41-44*	pJL37-Phsp60	*Hpa*I	Halo	PCR: JCV61, JCV62	Kan^R	oriM	
pJV61	Che9c gene *60*	pLAM12	*Hpa*I	pJV23	PCR: JCV47, JCV77	Kan^R	oriM	Pacet
pJV62	Che9c gene *61*	pLAM12	*Nde*I, *Nhe*I	pJV52	*Nde*I, *Nhe*I	Kan^R	oriM	Pacet
pJV63	Che9c genes *60-61*	pLAM12	*Nde*I, *Nhe*I	pJV53	*Nde*I, *Nhe*I	Kan^R	oriM	Pacet
pJV64	*Msmeg leuD* , R	pSJ25Hyg	*Hind*III, bl.	*M. smegmatis*	PCR: JCV113, JCV118	Hyg^R	Bxb1 attP/Int	
pJV64opal	*leuD* R15* R16* opal	pJV64	NA	NA	SDM: JCV132, JCV133	Hyg^R	Bxb1 attP/Int	
pJV64amber	*leuD* K31* R32* amber	pJV64	NA	NA	SDM: JCV134, JCV135	Hyg^R	Bxb1 attP/Int	
pJV67	*Msmeg recB* upstream homol.	pYUB854	*Bgl*II, *Xho*I	*M. smegmatis*	PCR: JCV136, JCV137	Hyg^R		γδ res sites
pJV68	*Msmeg recB* downstream homol; final recB KO plasmid	pYUB854	*Afl*II, *Xba*I	*M. smegmatis*	PCR: JCV138, JCV139	Hyg^R		γδ res sites
pJV69	pYUB854: Hyg - res sites	pYUB854	*Nhe*I, *Xba*I	pMsgroEL1 KO	PCR: JCV181, JCV182	Hyg^R		
pJV70	*groEL* KO: Hyg - res sites	pMsgroEL 1KO	*Nhe*I, *Xba*I	pMsgroEL1 KO	PCR: JCV181, JCV182	Hyg^R		
pJV71	MtbgroEL KO: Hyg - res sites	pMtbgroE L1KO	*Nhe*I, *Xba*I	pMsgroEL1 KO	PCR: JCV181, JCV182	Hyg^R		
pJV72	Che9c gene *61*	pJL37	*Hpa*I	pJV23	PCR: JCV49, JCV78	Kan^R	oriM	Phsp60
pJV73	Hyg^R cassette, F	pJL37	*Spe*I	pMsgroEL1 KO	PCR: JCV181, JCV183	Kan^R, Hyg^R	oriM	Phsp60
pJV74	Hyg^R cassette, F	pLAM12	*Spe*I	pMsgroEL1 KO	PCR: JCV181, JCV183	Kan^R, Hyg^R	oriM	Pacet
pJV75	Hyg^R cassette, F	pJV52	*Spe*I	pMsgroEL1 KO	PCR: JCV181, JCV183	Kan^R, Hyg^R	oriM	Pacet
pJV76	Hyg^R cassette, F	pJV53	*Spe*I	pMsgroEL1 KO	PCR: JCV181, JCV183	Kan^R, Hyg^R	oriM	Pacet
pJV77	Hyg^R cassette, F	pJV61	*Spe*I	pMsgroEL1 KO	PCR: JCV181,	Kan^R, Hyg^R	oriM	Pacet

pJV78	Hyg^R cassette, F	pJV72	*Spe*I	pMsgroEL1 KO	JCV183 PCR: JCV181, JCV183	Kan^R, Hyg^R	oriM	Phsp60
pJV73 amber	Hyg D15* D16* amber	pJV73	NA	NA	SDM: JCV184, JCV185	Kan^R	oriM	Phsp60; Hyg^S
pJV74 amber	Hyg D15* D16* amber	pJV74	NA	NA	SDM: JCV184, JCV185	Kan^R	oriM	Pacet; Hyg^S
pJV75 amber	Hyg D15* D16* amber	pJV75	NA	NA	SDM: JCV184, JCV185	Kan^R	oriM	Pacet; Hyg^S
pJV76 amber	Hyg D15* D16* amber	pJV76	NA	NA	SDM: JCV184, JCV185	Kan^R	oriM	Pacet; Hyg^S
pJV77 amber	Hyg D15* D16* amber	pJV77	NA	NA	SDM: JCV184, JCV185	Kan^R	oriM	Pacet; Hyg^S
pJV78 amber	Hyg D15* D16* amber	pJV78	NA	NA	SDM: JCV184, JCV185	Kan^R	oriM	Phsp60; Hyg^S
pJV73 opal	Hyg D15* D16* opal	pJV73	NA	NA	SDM: JCV186, JCV187	Kan^R	oriM	Phsp60; Hyg^S
pJV74 opal	Hyg D15* D16* opal	pJV74	NA	NA	SDM: JCV186, JCV187	Kan^R	oriM	Pacet; Hyg^S
pJV75 opal	Hyg D15* D16* opal	pJV75	NA	NA	SDM: JCV186, JCV187	Kan^R	oriM	Pacet; Hyg^S
pJV76 opal	Hyg D15* D16* opal	pJV76	NA	NA	SDM: JCV186, JCV187	Kan^R	oriM	Pacet; Hyg^S
pJV77 opal	Hyg D15* D16* opal	pJV77	NA	NA	SDM: JCV186, JCV187	Kan^R	oriM	Pacet; Hyg^S
pJV78 opal	Hyg D15* D16* opal	pJV78	NA	NA	SDM: JCV186, JCV187	Kan^R	oriM	Phsp60; Hyg^S
pJV79	Bxb1 attL(CT)-50, F	pMOSBlu eHyg	*Nde*I, bl.	attL(CT)-50mer, attL(CT)-50mer AP	NA	Hyg^R, Cb^R		
pJV80	Bxb1 attL(CT)-50, R	pJV69	*Xba*I, bl.	attL(CT)-50mer, attL(CT)-50mer AP	NA	Hyg^R		
pJV81	Bxb1 attR(CT)-50, F	pJV79	*Sma*I	attR(CT)-50mer, attR(CT)-50mer AP	NA	Hyg^R, Cb^R		Bxb1 attL/attR(CT) flanking Hyg^R
pJV82	Bxb1 attR(CT)-50, R	pJV80	*Nhe*I, bl.	attR(CT)-50mer, attR(CT)-50mer AP	NA	Hyg^R		
pJV83	*marinum* ATCC927 *recA*, F	pMOS Blue	*Eco*RV	*M. marinum* ATCC	PCR: JCV194, JCV197	Cb^R		
pJV84	*marinum* ATCC927	pYUB854	*Xba*I, bl.	*M. marinum* ATCC	PCR: JCV194,	Hyg^R		γδ res sites

	*recA*upstream homol., R				JCV195			
pJV85	*marinum recA* downstream homol; final KO plasmid, R	pJV84	*Nde*I	*M. marinum* ATCC	PCR: JCV196, JCV197	Hyg^R		γδ res sites
pJV86	Hyg^R cassette, R	pSJ25B	*Dra*I	pMOSBlue Hyg	PCR: AB01, AB02	Hyg^R	Bxb1 attP/Int	
pJV87	Hyg^R cassette, F	pSJ25B	*Dra*I	pMOSBlue Hyg	PCR: AB01, AB02	Hyg^R	Bxb1 attP/Int	
pJV88	L5 attP/Int cassette, R	pMOSBlueHyg	*Dra*I	pMH94	SalI, bl.	Hyg^R	L5 attP/Int	Interr. lacZ
pJV89A	$Gent^R$ cassette, F	pJV86	*Xmn*I	pJV43	*BamH*I, bl.	Hyg^R, $Gent^R$	Bxb1 attP/Int	
pJV89B	$Gent^R$ cassette, R	pJV86	*Xmn*I	pJV43	*BamH*I, bl.	Hyg^R, $Gent^R$	Bxb1 attP/Int	
pJV89 amber	Hyg D15* D16* amber	pJV89A	NA	NA	SDM: JCV184, JCV185	$Gent^R$	Bxb1 attP/Int	Hyg^S
pJV90A	SacB cassette, F	pJV86	*Xmn*I	pAVN30	PCR: sacBF/R	Hyg^R	Bxb1 attP/Int	sacB
pJV91A	$Gent^R$ cassette, F	pJV87	*Xmn*I	pJV43	*BamH*I, bl.	Hyg^R, $Gent^R$	Bxb1 attP/Int	
pJV91B	$Gent^R$ cassette, R	pJV87	*Xmn*I	pJV43	*BamH*I, bl.	Hyg^R, $Gent^R$	Bxb1 attP/Int	
pJV91 amber	Hyg D15* D16* amber	pJV91A	NA	NA	SDM: JCV184, JCV185	$Gent^R$	Bxb1 attP/Int	Hyg^S
pJV92A	$Gent^R$ cassette, F	pJV39	*Sma*I	pJV43	*BamH*I, bl.	Hyg^R, $Gent^R$	L5 attP/Int	
pJV92B	$Gent^R$ cassette, R	pJV39	*Sma*I	pJV43	*BamH*I, bl.	Hyg^R, $Gent^R$	L5 attP/Int	
pJV92 amber	Hyg D15* D16* amber	pJV92A	NA	NA	SDM: JCV184, JCV185	$Gent^R$	L5 attP/Int	Hyg^S
pJV93A	SacB cassette, F	pJV39	*Sma*I	pAVN30	PCR: sacBF/R	Hyg^R	L5 attP/Int	sacB
pJV93B	SacB cassette, R	pJV39	*Sma*I	pAVN30	PCR: sacBF/R	Hyg^R	L5 attP/Int	sacB
pJV94A	$Gent^R$ cassette, F	pJV88	*Sma*I	pJV43	*BamH*I, bl.	Hyg^R, $Gent^R$	L5 attP/Int	
pJV94 amber	Hyg D15* D16* amber	pJV94A	NA	NA	SDM: JCV184, JCV185	$Gent^R$	L5 attP/Int	Hyg^S
pJV95B	SacB cassette, R	pJV88	*Sma*I	pAVN30	PCR: sacBF/R	Hyg^R	L5 attP/Int	sacB
pJV96	Pacet	pJV44	*Nde*I, *Xba*I	pLAM12	*Nde*I, *Xba*I	$Gent^R$	oriM	Pacet
pJV97	Lambda *gam*	pJV44	*Hpa*I	Lambda	PCR: JCV145, JCV146	$Gent^R$	oriM	Phsp60
pJV98	Lambda *gam*	pJV53	*Nhe*I, bl.	Lambda	PCR: JCV145, JCV146	$Gent^R$	oriM	Pacet
pJV99	Lambda *gam*	pJV96	*Hpa*I	Lambda	PCR: JCV145, JCV146	$Gent^R$	oriM	Pacet
pJV100	*M. marinum* M strain recA KO	pYUB854	*Bgl*II, *Nco*I, *Afl*II, *Xba*I	*M. marinum* M strain	PCR: JCV208-211	Hyg^R		γδ res sites

pJV101	*M. smegmatis recD* KO	pYUB854	BlgII, *Hind*III, *Afl*II, *Xba*I	*M. smegmatis*	PCR: JCV212-215	Hyg^R		γδ res sites
pJV102	*attL-hyg*^{R}*-attR*	pJV39	*Xba*I, *Kpn*I	pJV81	*Xba*I, *Kpn*I	Hyg^R	L5 attP/Int	Bxb1 attL-INT/ attR(CT) flanking Hyg^R
pJV103	Halo gene *43*	pLAM12	*Hpa*I	pJV57	PCR: JCV53, JCV226	Kan^R	oriM	Pacet
pJV104	*E. coli recT*	pLAM12	*Hpa*I	pRDK557	PCR: JCV230, JCV237	Kan^R	oriM	Pacet
pJV105	Lambda *bet*	pLAM12	*Hpa*I	Lambda DNA	PCR: JCV227, JCV228	Kan^R	oriM	Pacet
pJV106	Halo gene *43*	pLAM12	*Nde*I, *Eco*RI	pJV57	PCR: JCV53, JCV226	Kan^R	oriM	Pacet
pJV107	*E. coli recT*	pLAM12	*Nde*I, *Eco*RI, bl.	pRDK557	PCR: JCV229, JCV237	Kan^R	oriM	Pacet
pJV108	Lambda *bet*	pLAM12	*Nde*I, *Eco*RI	Lambda DNA	PCR: JCV227, JCV228	Kan^R	oriM	Pacet
pJV109	Bxb1 attL (CT)-INT	pJV81	*Spe*I, *Hpa*I, bl.	Bxb1 lysogen	PCR: attL(CT)-50, JCV243	Hyg^R, Cb^R		Bxb1 attL-INT/ attR(CT) flanking Hyg^R
pJV110	Bxb1 attL (CT)-INT	pJV102	*Spe*I, *Hpa*I, bl.	Bxb1 lysogen	PCR: attL(CT)-50, JCV243	Hyg^R	L5 attP/Int	Bxb1 attL-INT/ attR(CT) flanking Hyg^R
pJV111	*Msmeg katG* (6384) KO	pYUB854	*Afl*II, *Xba*I, *Bgl*I, *Hind*III	*M. smegmatis*	PCR: JCV249-252	Hyg^R		γδ res sites
pJV112	Phsp60-lacZ	pJV45F	*Hind*III, *Xba*I	pMV261-lac	*Hind*III, *Xba*I	$Gent^R$	L5 attP/Int	lacZ, Interr. Hyg^R
pJV113R	sacB-new (no Phsp60), R	pJV53	*Nhe*I	pBRL301	PCR: JCV288, JCV289	Kan^R	oriM	Pacet, Che9c 60-61
pJV114F	sacB-new (no Phsp60), F	pJV62	*Nhe*I	pBRL301	PCR: JCV288, JCV289	Kan^R	oriM	Pacet, Che9c 61
pJV114R	sacB-new (no Phsp60), R	pJV62	*Nhe*I	pBRL301	PCR: JCV288, JCV289	Kan^R	oriM	Pacet, Che9c 61
pJV115F	sacB-new (no Phsp60), F	pJV75 amber	*Nhe*I	pBRL301	PCR: JCV288, JCV289	Kan^R	oriM	Pacet, Che9c 61, Hyg^S
pJV115R	sacB-new (no Phsp60), R	pJV75 amber	*Nhe*I	pBRL301	PCR: JCV288, JCV289	Kan^R	oriM	Pacet, Che9c 61, Hyg^S
pJV116	Giles gene *53*	pLAM12	*Nde*I, *Eco*RI	Giles	PCR: JCV338, JCV339	Kan^R	oriM	Pacet
pJV117	Halo gene *43*, alternative	pLAM12	*Nde*I, *Eco*RI	Halo	PCR: JCV379, JCV226	Kan^R	oriM	Pacet

	upstream start codon							
pJV118	tRN A amber suppressor	pSE100	*BamH*I, *Hind*III	pLT193B-B	PCR: JCV290, JCV291	Hyg^R	oriM	Tet operator
pJV119	$Gent^R$ cassette	pJV39	XmaI, *Spe*I	pJV44	PCR: JCV292, JCV293	$Gent^R$	L5 attP/Int	Interr. lacZ
pJV120	Pmyc1TetO p-tRNA amber suppressor, whole cassette 1, R	pJV119	*Sma*I	pLT193B-B	PCR: JCV333, JCV334, JCV346, JCV45	$Gent^R$	L5 attP/Int	Interr. lacZ
pJV121	Pmyc1TetO p-tRNA amber suppressor, gene 9	pJV119	*Sma*I	pLT193B-B	PCR: JCV334, JCV335, JCV346, JCV345	$Gent^R$	L5 attP/Int	Interr. lacZ
pJV122	sacB-new (no Phsp60), F	pJV69	*Nhe*I	pBRL301	PCR: JCV288, JCV289	Hyg^R		
pJV123	Che9c genes *60-61*	pTX-2MIX	*Pml*I	pJV23	PCR: JCV47, JCV78	Cb^R		TetR; Tet operator
pJV124	Che9c gene *61*	pTX-2MIX	*Pml*I	pJV23	PCR: JCV49, JCV78	Cb^R		TetR; Tet operator
pJV125	Msmeg *leuD* KO plasmid	pJV122F	*Afl*II, *Xba*I, *Hind*III, *Xho*I	*M. smegmatis*	PCR: Becky1+6, Becky3+8	Hyg^R		
pJV126	Phsp60-sacB (new), R	pJV53	*Nhe*I	pBRL301	*Xba*I	Kan^R	oriM	Pacet, Che9c 60-61
pJV127	Phsp60-sacB (new), R	pJV62	*Nhe*I	pBRL301	*Xba*I	Kan^R	oriM	Pacet, Che9c 61
pJV128	Phsp60-sacB (new), R	pJV75 amber	*Nhe*I	pBRL301	*Xba*I	Kan^R	oriM	Pacet, Che9c 61, Hyg^S
pJV129	Phsp60-sacB (new), R	pJV76 amber	*Nhe*I	pBRL301	*Xba*I	Kan^R	oriM	Pacet, Che9c 60-61, Hyg^S
pJV130	Hyg^R cassette, F	pMV261-lac	*Nhe*I, *Spe*I,	pMsgroEL1 KO	PCR: JCV181, JCV183	Hyg^R	oriM	Phsp60-lacZ
pJV131 R	TetR 1.7 (mutant), R	pJV75 opal	*Nhe*I	pTEK-4SOX	*Spe*I	Kan^R	oriM	Pacet, Che9c 61, Hyg^S
pJV132 R	TetR 1.7 (mutant), R	pJV127	*Spe*I	pTEK-4SOX	*Spe*I	Kan^R	oriM	Pacet, Che9c 61, SacB, Hyg^S
pJV133 R	TetR, R	pJV75 opal	*Nhe*I	pTX-2MIX	PCR: JCV363, JCV364	Kan^R	oriM	Pacet, Che9c 61, Hyg^S
pJV134	TetR, R	pJV127	*Spe*I	pTX-2MIX	PCR: JCV363, JCV364	Kan^R	oriM	Pacet, Che9c 61, SacB, Hyg^S
pJV135 F	Phsp60-sacB (new), F	pJV119	*Xba*I	pBRL301	*Xba*I	$Gent^R$	L5 attP/Int	Interr. lacZ

pJV135R	Phsp60-sacB (new), R	pJV119	*Xba*I	pBRL301	*Xba*I	Gent^R	L5 attP/Int	Interr. lacZ
pJV136F	Gent^R cassette, F	pJV39	*Sma*I, *Hpa*I	pJV44	PCR: JCV292, JCV293	Gent^R	L5 attP/Int	Interr. lacZ
pJV136R	Gent^R cassette, R	pJV39	*Sma*I, *Hpa*I	pJV44	PCR: JCV292, JCV293	Gent^R	L5 attP/Int	Interr. lacZ
pJV137	Pmyc1TetOp-tRNA amber suppressor	pJV119	*Hind*III, *Spe*I	pJV118	*Hind*III, *Spe*I	Gent^R	L5 attP/Int	Interr. lacZ
pJV138	Hyg^R cassette, R	pJV53	*Nhe*I, *Spe*I	pMsgroEL1 KO	PCR: JCV181, JCV183	Hyg^R	oriM	Pacet, Che9c 60-61
pJV139	tRNA amber suppressor (whole cassette 2)	pSE100	*Eco*RV	pLT193B-B	PCR: JCV380, JCV381	Hyg^R	oriM	Tet operator
pJV140	Phsp60RBS-lacZ	pSE100	*Eco*RV	pMV261-lac	PCR: JCV384, JCV10	Hyg^R	oriM	Tet operator
pJV141F	Pmyc1TetOp-tRNA amber suppressor, whole cassette 2, F	pJV136F	*Xba*I, bl.	pJV129	*Spe*I, ClaI, bl.	Gent^R	L5 attP/Int	Interr. lacZ
pJV141R	Pmyc1TetOp-tRNA amber suppressor, whole cassette 2, R	pJV136F	*Xba*I, bl.	pJV129	*Spe*I, ClaI, bl.	Gent^R	L5 attP/Int	Interr. lacZ
pJV142	tRNA amber suppressor cassette	pMPambar	*Dra*I, *Xho*I, bl.	pLT193B-B	PCR: JCV380, JCV381	Gent^R	L5 attP/Int	TetR repressor, Tet operator-tRNA
pJV143	Lambda *gam*	pJV44	*Nde*I	NA	NA	Gent^R	oriM	Phsp60
pJV144	Lambda *gam*	pJV44	*Nde*I	NA	NA	Gent^R	oriM	Pacet
pJV145	Giles gene *53*	pLAM12	*Hpa*I	Giles	PCR: JCV338, JCV339	Kan^R	oriM	Pacet
pJV146	Phsp60RBS-lacZ	pMPambar	*Dra*I, *Xho*I, bl.	pMV261-lac	PCR: JCV10, JCV384	Gent^R	L5 attP/Int	TetR repressor, Tet operator-lacZ
pJV148	Hyg^R cassette	pTTP1B	*Hind*III, bl.	pMsgroEL1 KO	PCR: JCV181, JCV183	Hyg^R, Cb^R	Tweety attP/Int	
pJV149	Phsp60-sacB (new), R	pMsgroEL 1KO	*Nhe*I	pBRL301	*Xba*I	Hyg^R		groEL1 KO plasmid, sacB
pJV150F	Phsp60-sacB (new),	pJV69	*Nhe*I	pBRL301	*Xba*I	Hyg^R		Hyg^R, sacB KO

pJV150R	F Phsp60-sacB (new), R	pJV69	*Nhe*I	pBRL301	*Xba*I	Hyg^R	plasmid Hyg^R, sacB KO plasmid

6.2.3 Cloning procedures

6.2.3.1 Preparation of the insert and vector for plasmid constructions

Plasmid cloning was performed using DNA inserts generated either by PCR-amplification or restriction digest of the parental insert source as described in Table 15. PCR reactions were set up as described below (PCR), and the products were cleaned up using the QIAquick PCR cleanup protocol (QIAGEN), eluting in 50 µl EB buffer. Restriction digests of either plasmid DNA or PCR DNA were used to obtain the vector backbone or desired insert. Restriction enzymes were from NEB exclusively, and digests were performed according to manufacturer's instructions for preferred buffers and reaction conditions, typically in 50 µl reaction volumes. Restriction digests were incubated at the required temperature for 2 hr for plasmid DNAs and 4 hr for PCR products. Following digestion, the vector backbone digest was heat-killed (if possible) and treated with 1 µl calf intestinal phosphatase (CIP; NEB) to remove the 5′ phosphate and prevent ligation of the vector to itself. The digest reactions were then run on 0.8% agarose gels, the bands containing the desired DNA fragments were extracted, and these were cleaned up using the QIAquick gel extraction protocol (QIAGEN). The DNA was quantified by analyzing 2 µl on an agarose gel and comparing to the GC #1 quantitative DNA ladder (GeneChoice) using Quantity One 4.6 software.

In cases in which a PCR product was cloned into a blunt site in the parental plasmid, the PCR product was treated with T4 polynucleotide kinase (Roche) and the accompanying buffer

directly in the PCR reaction (without cleanup) for 1 hr at 37°C. Plasmids and inserts that required blunt ends were treated with the Klenow enzyme (NEB). The DNA digest was first heat-killed (if possible), and subsequently 0.5 μM dNTPs, 1X buffer were added, and 0.5 μl of Klenow in a total reaction volume of 60 μl. The reactions were incubated at room temperature for 15 mins, and these were immediately cleaned up using the QIAquick gel extraction protocol for enzymatic cleanup (QIAGEN) or run on an agarose gel to be gel extracted.

6.2.3.2 Ligations and transformations

Ligations were performed with the Fast-link DNA ligase enzyme (Epicentre) in 15 μl volumes using 1.5 μl ATP for sticky-end ligations and 0.75 μl ATP for blunt-end ligations. These were incubated at room temperature for 2 hr or longer and heat-killed at 75°C for 15 mins. The heat-killed ligations were transformed into *E. coli* GC5 cells as described below (*E. coli* transformations). All plasmids that were constructed were checked by restriction digest with two different enzymes, and the insert was sequenced.

6.3 PCR

Primers were designed typically with melting temperatures at or above 60°C (if possible) to simplify amplification from the high-G+C% mycobacterial or mycobacteriophage templates. If restriction sites were used, these were engineered in the center of the oligonucleotide by changing nucleotides when required, with perfect homology to the template flanking the site. Oligonucleotides were synthesized as described above (DNA substrates) and resuspended in TE

buffer to 100 μM stock solutions upon receipt of the lyophilized pellet of DNA. Working solutions were 10 μM, and all stock solutions were stored at -20°C.

PCR reactions were performed as described [7], typically using *Pfu* polymerase (Stratagene) and when necessary, *Pfu* Turbo (for extended length targets longer than 4 kbp). Reactions (50 μl) contained template DNA (5-10 ng), 0.5 μM primers, 0.2 mM dNTPs, 1X buffer, 0-5% DMSO (typically 5%), and 1-5 U polymerase. Cycling conditions were set with an initial 95°C denaturation for 5 min, and 25 cycles of denaturation (95°C for 30 sec), annealing (varying temperature, 30 sec), and extension (72°C, varying length). Annealing temperatures used were 2°C lower than the lowest primer melting temperature, and the extension was equal to approximately 1 min per 1 kbp of desired product. A final extension at 72°C for 7 mins was used, followed by cooling to 4°C. The 72°C extension temperature was used only for *Pfu* polymerase; it was adjusted to 68°C for all *Taq* polymerases.

6.3.1 Colony PCR

PCRs used as a screening method (see section 7.10.4) typically were performed with *Taq* DNA polymerase (NEB). Colonies were resuspended using sterile toothpicks in 200 μl dH_2O, vortexed vigorously, boiled for 5 min, and vortexed again, and a volume equal to 1/10 the total PCR reaction volume was used. Reaction volumes often were decreased to 25 μl for screening PCRs.

6.3.2 MAMA-PCR

Mismatch amplification mutation assay PCR (MAMA-PCR) [30,219] was used to identify mutant alleles – either point mutations or deletions – in a population consisting primarily of wild

type alleles. Primers (~20 nt) were designed in which the 3′ ultimate base of the primer matched the mutant sequence. For point mutations, this was the most 3′ of the mutated bases; for deletions, this was at the junction of the new deletion allele locus. Typically, the penultimate base was also changed such that neither base would anneal at a wild type locus, and only the ultimate base would anneal at a mutant locus (Figure X). Reactions were performed using Platinum *Taq* High Fidelity DNA Polymerase (Invitrogen), in which 100 μl of culture or resuspended colonies were used for colony PCR as described above. Reaction conditions were as described above for basic PCR with 2 mM $MgSO_4$ added to increase the fidelity of the polymerase.

6.3.3 Reverse transcription-PCR

Reverse transcription-PCR (RT-PCR) was performed essentially as described [228]. RNA was extracted and purified from *M. smegmatis* cultures using the RiboPure-Bacteria kit (Ambion). Each sample was disrupted with the mini bead-beater (30 sec) following an 8 min vortexing on the vortex adapter. RNA samples were treated with DNaseI, as well as RNaseOUT to kill any contaminating RNase enzymes; DNase Inactivation reagent was used following DNase treatment. RNA aliquots were analyzed on agarose gels, quantified and stored at -80°C. RT-PCR reactions were performed with the Qiagen OneStep RT-PCR kit (Qiagen) with equal amounts of RNA and analyzed on agarose gels. Control PCR reactions using *Pfu* polymerase (Stratagene) were used as controls to assess the presence of contaminating DNA in the RNA samples.

6.3.4 Sequencing

Sequencing of plasmids or PCRs was done either through the DNA synthesis facility of the University of Pittsburgh or GeneWiz company. Typically 5-8 μl of plasmid miniprep DNA was used (~800 ng), and either 3.2 pmols (DNA synthesis facility) or 8 pmols (Genewiz) of primer were used. Sequencing primers were designed 100 bp upstream of the region to be sequenced.

6.3.5 Site-directed mutagenesis (SDM)

The QuikChange site-directed mutagenesis (and sometimes the XL kit; Stratagene) was used for all experiments. Primers were typically PAGE-purified, and PCR conditions were used as recommended by the manufacturer.

6.4 DNA SUBSTRATES

All oligonucleotides (Table 16) were purchased from Integrated DNA Technologies, and when required (lengths exceeding 30 nt), the oligonucleotides were PAGE-purified. When necessary, oligonucleotides were radiolabeled with [γ-^{32}P]-ATP (Perkin-Elmer) and T4 polynucleotide kinase (Roche) for 30 min at 37°C and purified on ProbeQuant G-50 Micro Columns (Amersham) to remove unincorporated label. When short (50 – 100 bp) dsDNA substrates were required, complementary oligonucleotides were annealed at 0.4 μM in annealing buffer (50 mM Tris, pH 7.5 and 100 mM NaCl) by incubating for 5 min at 95°C in a water bath and slow cooling overnight to room temperature.

Table 16: Oligonucleotides.

Primer Name	Sequence
TM413069-13089 1089F	CAAGGCTATCGAGGACAAGCA
TM4414137-14157 1089R	GGGTGGCAGTAATACCACTTG
TM41444-1464 880F	ATGCGTAAAGCGTTGGGCGAT
TM42323-2303 880R	TCGCCAGTTCCTTGACTTCGT
TM430497-30517 699F	CCTGCTGTGCACCAAGTGCTT
TM431195-31175 699R	TCCTGCACGACTCGATGTTCT
TM446441-46461 1328F	GCGTGTTGACAGCTCAACAGT
TM447748-47769 1328R	GTCATGTGGTTGGTCATCTCG
1512-1536F	ACGCTCAGTCGAACGAAAACTCACG
1624-1648R	AGCTTCGTGGATCCAGATATCCTGC
pYUB854 1442-1466F	TGCAAGCAGCAGATTACGCGCAGAA
pYUB854 1775-1751R	TCAGATATCGGACAAGCAGTGTCTG
pYUB854 509-533F	ATGATCGTGCTCCTGTCGTTGAGGA
pYUB854 1069-1093R	TGAGCTATGAGAAAGCGCCACGCTT
pJL37F	ACTGCGCCCGGCCAGCGTAAGTAGC
pJL37R	ATCAGAGATTTTGAGACACAACGTGG
LJM20	CCGCAGTTGTTCTCGCATACCCCATC
LJM23	CGGACGGTTGCTAGCACGCGCACCAT
SACB-F	GGACATCCTGAGCTTGCTAGAGGA
SACB-R	CTCGACGACCTGCAGGATCG
M13-50MER	AAACAGCTATGACCATGATTACGAATTCGAGCTCGGTACCCGGGGATCCT
M13-50MER-AP	AGGATCCCCGGGTACCGAGCTCGAATTCGTAATCATGGTCATAGCTGTTT
DJ20	CGTAGGAATCATCCGAATCA
DJ76	CAGAATTCCTGGTCGTTCCGCAGGCTCGCGTAGGAATCATCCGAATCAATACGGTCGAG AAGTAACAGGGATTCTT
AB01	GTCGACTCTAGAGGATCTACTAGTC
AB02	GTAAAACGCTAGCCAGTGAATTCGAG
Bxb1 attB-50	TCGGCCGGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCCGGGC
Bxb1 attB-50-AP	GCCCGGATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCCGA
Bxb1 attB-100	TGGCCGTGGCCGTGCTCGTCCTCGTCGGCCGGCTTGTCGACGACGGCGGTCTCCGTCGT CAGGATCATCCGGGCCACCGAGGCGGCGTTGAGAACAGC
Bxb1 attB-100-AP	GCTGTTCTCAACGCCGCCTCGGTGGCCCGGATGATCCTGACGACGGAGACCGCCGTCGT CGACAAGCCGGCCGACGAGGACGAGCACGGCCACGGCCA
JCV01	GCAAGGTCGTCACCGAGCGGTTCAA
JCV02	TTCTCGACGGCCTTCACGATCACCT
JCV03	CTCGAAATGCGTCACCTCGTACAC
JCV04	GGAGACAGGTGCATATGACAACCGA
JCV05	CAGCAGGTCGACGCGGTAGTGCCTC
JCV06	GACTTGATCAGAAGCTTGATGCGGT
JCV07	GAACTCCGTTGTAGTGCTTGTGGTG
JCV08	TCGGTTGCACTACGTGTACTGTGAG
JCV09	TGCCAATGAATGCTCTGACCGATGA
JCV10	AACTACGTCGGCATTCATAAGCTTC
JCV11	CATTCGCCATTCAGGCTGCGCAACT
JCV12	CATTAGGCACCCCAGGCTTTACACT
JCV13	AGCACGAGTCGCTGTTCGAGCTACC
JCV14	ACTACCTGCGAAAACACGTCGATAC
JCV15	GCGCGGTAGGCCTTTCGCGGTTCTG
JCV16	ACAGCGATGCTCTAGATGGCGAT
JCV17	GCCGAGCTCGAGACCCAGCTGACCA
JCV18	GAACGCAGGCTGCGCAGATCTACCCGC
JCV19	CTGACGATCGAATCGAGACGGAGAA
JCV20	CGCTGGCTGGCATCTCCAGGTCGA
JCV21	GGGCGAAAAACCGGATATCAGGGCGATGGCCCAC

JCV22	GTGGGCCATCGCCCTGATATCCGGTTTTTCGCCC
JCV23	GACCAAAATCCCTTCACGTGAGTTTTCGTTCCAC
JCV24	GTGGAACGAAAACTCACGTGAAGGGATTTTGGTC
JCV25	TACGCGCCGTGTAAGGGCACGCAGA
JCV26	GGCAGCGAGGACAACTTGAGCCGTC
JCV27	GGTTACGATGCGCCCATCTACACCA
JCV28	CACTCGCTTTAATGATGATTTCAGC
JCV29	CTGCATGGTCAGGTCATGGATGAGC
JCV30	TGCGCCAAGCTTCCTGCTGAACATC
JCV31	CATACACGGTGCCTGACTGCGTTAG
JCV32	GTCAGATGCGGGATGGCGTGGGACG
JCV33	CAAAACAGGCGGCAGTAAGGCGGTCG
JCV34	GAGTTGGTAGCTCTTGATCCGGCAAAC
JCV35	GACCCCGAGGTGCACGGGCTGAAGGC
JCV36	CGGCTCTACGCCGACTGCATGGAATC
JCV37	CTACCAGGGCATCGTCAAACTGTTC
JCV38	CAACGACGCCGATTGGTTCAAGTTCC
JCV39	CTTCAGCAGAGCGCAGATATCAAATACTGTCCTTC
JCV40	GAAGGACAGTATTTGATATCTGCGCTCTGCTGAAG
JCV41	CGTAGTTAGGCCACCACGTGAAGAACTCTGTAGCACC
JCV42	GGTGCTACAGAGTTCTTCACGTGGTGGCCTAACTACG
JCV43	GTGGTTTGTTTGCCGGATCACGTGCTACCAACTCTTTTTCC
JCV44	GGAAAAAGAGTTGGTAGCACGTGATCCGGCAAACAAACCAC
JCV45	CTGGCTTCAGCAGAGCGCATATGCCAAATACTGTCCTTC
JCV46	GAAGGACAGTATTTGGCATATGCGCTCTGCTGAAGCCAG
JCV47	ACGAGATCGGCGGCCGCATATGAGT
JCV48	TGCTTGGTGACAGCACTCGAGGCCATA
JCV49	AAGGGGATTACACATATGGCTGAAA
JCV50	TGGTTCGCCGGGGAAGCTTTGCGTT
JCV51	TGATGAACGGGCCCCGCATATGAC
JCV52	TGTTTGGTGCCTCTCGAGTGCGGT
JCV53	TGAGGAAGGCACCAAACCATATGACC
JCV54	TGAATCAGTCCCAAAGCTTCTCGT
JCV55	GACGGCTCAAGTTGTCCTCGCTGCC
JCV56	TTTCGCTAAATACTGGCAGGCGTTT
JCV57	CACATGTTCTTTCCTGCGTTATCCC
JCV58	GCTAACTCACATTAATTGCGTTGCG
JCV59	GAATCAGTCCCAAAGCTTCTCGTTCCCTTCAC
JCV60	GAATCAGTCCCAGAATTCCTCGTTCCCTTCAC
JCV61	GTCGACGAACTGCTCGAACGACTCAT
JCV62	GTGAGTCGCGGAACATCGCGAAGCAC
JCV63	CTGCGCCGTTACCACCGCTGCGTTC
JCV64	GATATCGACCCAAGTACCGCCACCT
JCV65	CTCTCATGTATAGAGTGCTAGGTGGC
JCV66	GGTCGTGGTTTCGCCGCCAGGAGCGGA
JCV67	GTGACACCGTCATCTACAGCAAGTACGGC
JCV68	CGGTCTCGCAGGCCTCGAAAAACCGCTCAG
JCV69	GTCCGGGGCGGTACGCGTAGGCGTCTGAAAG
JCV70	CTCCGTCGTCAGGATCATCCGGGCCAC
JCV71	AGCGGCTCCCAGAATTCCTGGTCGTTC
JCV72	GCGAACTGCTCGCCTTCACCTTCCTGCAC
JCV73	GTCCGGGGCGGTACGCGTAGGCGTCTGAAAGAGACTTATGAGCAATCTAGGGGGATCC GATAAATCAATCTAAAG
JCV74	CTCCGTCGTCAGGATCATCCGGGCCACCGAGGCGGCGTTGAGAACAGGCAGGTCGACT CTAGAGGATCTACT
JCV75	TCGCGCCCACCACGACACCGAACGG
JCV76	GACGTCACCGTCGCGCAGTCGATCGT
JCV77	GTGCCTTGGGCGAATTCTGCTTGGT
JCV78	GCAGATCAACCACCGCGTCGGAATTCGC
JCV79	GAAATTCCGCTGGGAGATGGACAACA
JCV80	CGACCAGTAGATCGATGTACCAGG
JCV81	GTGTCAACGGCGAGTGCTACCTC

JCV82	CGAGCGCCTGCGACTGTGACAGG
JCV83	GTCCCGAAGAACGATGGCATTTACG
JCV84	CACCAGCAGCAGGCGTTCTACGAAG
JCV85	GATTCCGCGGGAAGCCGGAAGAGGCT
JCV86	GTGAAGAAGTCGGTCGTCGGAAACA
JCV87	GAAAACCCGGATGCGGCAAGCACTC
JCV88	CGAACGCGCCGAAATCTATCTGTCC
JCV89	CACGGCCTCGACGGGTCCGTCGACCA
JCV90	GCGTCCACCTGCTGTTTTGCCTTCA
JCV91	GTGCGGCGCAGCACCATGACGTTGT
JCV92	CTCGTGCTCGGCCGGGCACGCAGGCG
JCV93	GTTCCCACTCGGAGGCAGGCAGA
JCV94	CGGCGACGTATCGGCGCACCTCAT
JCV95	GGGGATCCGATAAATCAATCTAAAG
JCV96	GCAGGTCGACTCTAGAGGATCTACT
JCV97	CTTATGAGCAATCTAGGGGGATCCGATAAATCAATCTAAAG
JCV98	CGGCGTTGAGAACAGGCAGGTCGACTCTAGAGGATCTACT
JCV99	GGATCGATGTCGACTGCCAGGCATCA
JCV100	CCAGTGAATTCGAGCTCGGTACCC
JCV101	CGCCAAGCTCTAATACGACTCACTA
JCV102	CGACGTTGTAAAACGACGGCCAGTGA
JCV103	GTCCGGGGCGGTACGCGTAGGCGTCTGAAAGAGACTTATGAGCAATCTAGGGATCGAT GTCGACTGCCAGGCATCA
JCV104	CTCCGTCGTCAGGATCATCCGGGCCACCGAGGCGGCGTTGAGAACAGGCCCAGTGAAT TCGAGCTCGGTACCC
JCV105	GGTCGCGACGAATCGAAAACGGTGCGA
JCV106	CAGCGTCGCCGCCGCGATGTTCTAT
JCV107	GCTTCCACCGGTTGCCTGCGCGTTCT
JCV108	GACGCGTTCGCCACACACCAGTGCGT
JCV109	GATCACCTCGTCGTCCACGCCCTT
JCV110	CACGGGCACAGCAGGATGAGGTC
JCV111	CAAGCGCGACGGAGACGGCCAGGGTA
JCV112	CACCTGGTGGATATCGAGTCCGGT
JCV113	CTGCCGCCCGGCGTGAGCGCCAAGGA
JCV114	GGACTGGTGGGCGCCGACTATCTGA
JCV115	GTGCACCAACGGCCGGATCGAGGAT
JCV116	GAGGTCGGCTTGACCTTCACGGTCT
JCV117	GAAAGCTCGCGTCCCCTGCAGATCTA
JCV118	CGCAGTTTCAGGCGAAGCTAGCGAA
JCV119	CATCGGTGTCCCGCTGCGGCGGTCCAA
JCV120	CGAGCCGCCTCGTAGGCCTCGATCTGA
JCV121	GTGTCCCGCTGCGGCGGTCCAACGTCGACACCGACCAGATCATCCCGGCGGGGCCTTA CGCTGCGGAAACTCGATCAGATCGAGGCCTACGAGGCGGCTC
JCV122	GAGCCGCCTCGTAGGCCTCGATCTGATCGAGTTTCCGCAGCGTAAGGCCCCGCCGGGAT GATCTGGTCGGTGTCGACGTTGGACCGCCGCAGCGGGACAC
JCV123	GACTCTAGAGGATCTACTAGTCATA
JCV124	GAATTCGAGCTCGGTACCCGGGGAT
JCV125	TACGACTCACTATAGGGAAAGCTTG
JCV126	CTGGATCCACGAAGCTTCCCATGGT
JCV127	AGAGCTCACCTAGGTATCTAGAAC
JCV128	GTGTCACCGGTGTTCGGTGACCGACT
JCV129	ACCATGGGAAGCTTCGTGGATCCAGCTCCAGCACCGTGGTGGTGTTCGGT
JCV130	GTTCTAGATACCTAGGTGAGCTCTGTATCTGTTCGTACGTGGCATGTGC
JCV131	CGTCAGCCCTTGCGAGAGCGCAA
JCV132	CATCCCGGCGGTCTACCTGTAGTAGGTGACCCGAACGGGTTTCG
JCV133	CGAAACCCGTTCGGGTCACCTACTACAGGTAGACCGCCGGGATG
JCV134	CCGGCATCGGTGTCCCGCTGTGATGATCCAACGTCGACACCGACC
JCV135	GGTCGGTGTCGACGTTGGATCATCACAGCGGGACACCGATGCCGG
JCV136	GTGTCACAGATCTTCGGTGACCGACT
JCV137	CGCTGGCCTCGAGCACCGTGGTGGTGTT
JCV138	GCGACGCATCTAGATGGTGTGCTGTAT
JCV139	CGTCAGCCCTTAAGAGAGCGCAA

JCV140	ACCATGGGAAGCTTCGTGGATCCAGCGCTGGCCTCCAGCACCGTGGTGGTGTT
JCV141	CTTCAGCAGATCCACCGCCTCGGT
JCV142	TGCCGAAATCCGGTCCGGCGACAA
JCV143	GATGCGCGACATCGCCGTGGACA
JCV144	ACGCCAGCTGGCGAAAGGGGGATGT
JCV145	CGTCGACGCTTATAAACATATGGATATT
JCV146	CGACACGTTCAGCAAGCTTCCCAG
JCV147	CACCAGTGAAGGGAACATATGACCA
JCV148	GAACATCGCGAAGCTTGCCGCGCAC
JCV149	GTCCGGGGCGGTACGCGTAGGCGTCTGAA
JCV150	GGTCTCCGTCGTCAGGATCATCCGGGC
JCV151	GTAATCACTCGTGTTCACCGCCCC
JCV152	TGGTGGCCGTGGCCGTGCTCGTCCT
JCV153	GAGGAGTACCTGATCCTGTCGGCCC
JCV154	CAAGCGGACCGGGGGTGTCGCGTGA
JCV155	ACACCGTCATCTACAGCAAGTACG
JCV156	TCGAAAAACCGCTCAGCGGCGCCCGCA
JCV157	CCCACGAAGGAGAAGCGTGATGGAGGCTTTCACCACTCACACCGGCATCGGTGTCCCG CTGCGGCGGTCCAA
JCV158	CGCCGAAGGCCCGGATCAGGCCGGGAGAGTGCGCGGTTTCCAGGCCGGACGAGCCGCC TCGTAGGCCTCGATCTG
JCV159	CGAACTGCTCGCCTTCACCTTCCT
JCV160	CTGTCCGCACCGCGGTCAGGCGTTG
JCV161	CAGCAACGCCAACAGCCGTGCCACGGT
JCV162	CCGTACGTCTCGAGGAATTCCTGCAGGATATCTGGATCCACGAAGCTTCCGGTCTCCGT CGTCAGGATCATCCGGGC
JCV163	TCGTTCATCCATAGTTGCCTGACTCCCAGTCCGTAATACGACTCACTTAAGTCCGGGGC GGTACGCGTAGGCGTCTGAA
JCV164	GGTCTCCGTCGTCAGGATCATCCGGGCCACCGAGGCGGCGTTGAGAACAGGACACGAC TTATCGCCACTGGCAGC
JCV165	GTCCGGGGCGGTACGCGTAGGCGTCTGAAAGAGACTTATGAGCAATCTAGGGAAACGA CAGGTGCTGAAAGCGAGCT
JCV166	CATGCGTTAGAGGTCGGTGAGCCCT
JCV167	CACCGTCATCTACAGCAAGTACG
JCV168	CTCATCTATCCGCCCGGGATAGCAT
JCV169	AAGTGGTAATTCGGACGGTTCCG
JCV170	GTGAGGAGTTTCATTCACCGTGATA
JCV171	GCTGGACGTTGCGGAGGGTGACA
JCV172	CTCACCGGTGAACGGGTGTTCGAT
JCV173	GTGCGGCGTGATCGCCGTCGGTAGC
JCV174	GATGCCAACGGGCCCGCCGGCACCA
JCV175	GCTGACCGGAGTTCAGTGCGCGTG
JCV176	CTGCGTAGAGGAGCCTGATGAGCAA
JCV177	CTTGCTGATGCCCGAACCCAGCGCGAT
JCV178	GAGACCGACCTGAAGAAGCGCAAGGA
JCV179	GAAAGGCCGGTGCGGTGAAGGTTTT
JCV180	CATGCACATCGGCGGGTTCGAGGATCT
JCV181	GATTTAGGATACATGCTAGCCACCT
JCV182	ACCGCAGCACCATCTAGAACGTCCC
JCV183	ACCGCAGCGCTAGCGAGAACGTCCC
JCV184	CGACCGTATTGATTCGTAGTAGTCCTACGCGAGCCTG
JCV185	CAGGCTCGCGTAGGACTACTACGAATCAATACGGTCG
JCV186	CGACCGTATTGATTCGTGATGATCCTACGCGAGCCTGC
JCV187	GCAGGCTCGCGTAGGATCATCACGAATCAATACGGTCG
JCV188	CTTCTCGACCGTATTGATTCGTAGTAGTCCTACGCGAGCCTGCGG
JCV189	CCGCAGGCTCGCGTAGGACTACTACGAATCAATACGGTCGAGAAG
JCV190	CCGTATTGATTCGGATGATTGATGAGCGAGCCTGCGGAACGACC
JCV191	GGTCGTTCCGCAGGCTCGCTCATCAATCATCCGAATCAATACGG
JCV192	GAGTTGGTAGCTCTTGATCCGGCA
JCV193	CTCCACGAGCTGCCCGTGGAGGACT
JCV194	GAACTACGCCCTGGCTCAAGACGCA
JCV195	CGGGATGATCGAAATCGGTTGCCGT

JCV196	GGCTTCATCCGGAAATCCGGTGCGT
JCV197	GCTTTTCGCGCCTGAGCTTGGTCCT
JCV198	CCGCTGTGACACAAGAATCCCTGTTACTTCTCGACCGTATTGATTCGGATGATTCCTAC GCGAGCCTGCGGAACGACCAGGAATTCTGGGAGCCGCTGGC
JCV199	GCCAGCGGCTCCCAGAATTCCTGGTCGTTCCGCAGGCTCGCGTAGGAATCATCCGAATC AATACGGTCGAGAAGTAACAGGGATTCTTGTGTCACAGCGG
JCV200	GACGAAAGGGCCTCGTGATACGCCTA
JCV201	CGAGCTCGGTACCCGGACATCCTGA
JCV202	CGATGTCGACTGCCAGGCATCAAAT
JCV203	GTTGGTGGGTGGCCGTGCATGTGAT
JCV204	TCAGGATGTCCGGGTACCGAGCTCGGGAAGCGGAAGTGCGCGCTGCGATA
JCV205	ATTTGATGCCTGGCAGTCGACATCGGTCAGCCGTTGCGCCACATGCACAT
JCV206	CTGCGTGGTGGACGGTTCGCACGGTTGT
JCV207	CATTCGCCATTCAGGCTGCGCAACT
JCV208	GAACTAAGATCTGGCTCAAGACGCA
JCV209	CGGGATGATCGACCATGGTTGCCGT
JCV210	GGCTTCATCTAGAAATCCGGTGCGT
JCV211	GCTTTTCGCGCTTAAGCTTGGTCCT
JCV212	GTTGGTGGGAGATCTTGCATGTGAT
JCV213	GGAAGCGGAAGCTTGCGCTGCGATA
JCV214	GTCAGCCGTCTAGACACATGCACAT
JCV215	CTGCGTGGTGGACGCTTAAGACGGTTGT
JCV216	GGATCACCGCCGAGATCGGTGAGGGCAACAAGATCGACGGTGTGGTGCACGCGATCGG GTTCATGCCGCAGAGCGGTATGGGCATCAACCCGTTCTTCGAC
JCV217	GTCGAAGAACGGGTTGATGCCCATACCGCTCTGCGGCATGAACCCGATCGCGTGCACC ACACCGTCGATCTTGTTGCCCTCACCGATCTCGGCGGTGATCC
JCV218	CAGCCCGCAGCGTCGCGGCGTGTGCACGCGCGTTTACACCACCACTCCGAGGAAGCCG AACTCGGCGCTCCGGAAGGTCGCGCGCGTGAAGCTGACCAGCC
JCV219	GGCTGGTCAGCTTCACGCGCGCGACCTTCCGGAGCGCCGAGTTCGGCTTCCTCGGAGTG GTGGTGTAAACGCGCGTGCACACGCCGCGACGCTGCGGGCTG
JCV220	TGTGGTCACCGACCAGATCGACTACCTGACCGCCGACGAGGAGGACCGCCATGTCGTG GCGCAGGCCAACTCGCCGACCGACGAGAACGGCCGCTTCACCG
JCV221	CGGTGAAGCGGCCGTTCTCGTCGGTCGGCGAGTTGGCCTGCGCCACGACATGGCGGTC CTCCTCGTCGGCGGTCAGGTAGTCGATCTGGTCGGTGACCACA
JCV222	TGTGGTCACCGACCAGATCGACTACCTGACCGCCGACGAGGAGGACCGCCGCGTCGTG GCGCAGGCCAACTCGCCGACCGACGAGAACGGCCGCTTCACCG
JCV223	CGGTGAAGCGGCCGTTCTCGTCGGTCGGCGAGTTGGCCTGCGCCACGACGCGGCGGTC CTCCTCGTCGGCGGTCAGGTAGTCGATCTGGTCGGTGACCACA
JCV224	GCTGGTGTAGTCGTGGCCGTTTCGAT
JCV225	CATTGTCGAAGCTGTTGGATGCGGA
JCV226	GCTGGTCCTGAATTCAGTCCCATGGT
JCV227	CAGAGGTATAAAACATATGAGTACTGCACT
JCV228	GCAGGAGAATTCCCGGTGTCATGCT
JCV229	TGACTAAGCAACCACCAATCGCAA
JCV230	GAATATGCAAATGACTAAGCAACCACC
JCV231	GACAGGCCTACTCGAAGGCAAGCGCA
JCV232	GCGCTCTTGGCGACGGTCATCCAGT
JCV233	GCCAACCATTCAGCAGCTGGTCCGCA
JCV234	GCTCTTCTCCTTCTTCGCGCCATAGC
JCV235	CCGGCCTCGAGGTCCGCGACGTGCA
JCV236	CACCAGCGGTGCCTCGCTGCGCACCA
JCV237	ACACCGCCAGGCTGAATTATTCCTCTG
JCV238	GATCACACCCGTGATCACAGCCCAATTCACCACTCCCGAAAGGAAATGCACACACAAC CACCTGGACGCCCAGGCCGGCCTCACAGCCGGCCTGGGCGTT
JCV239	AACGCCCAGGCCGGCTGTGAGGCCGGCCTGGGCGTCCAGGTGGTTGTGTGTGCATTTCC TTTCGGGAGTGGTGAATTGGGCTGTGATCACGGGTGTGATC
JCV240	CACAGCCCAATTCACCACTCCCGAAAGGAAATGCACACACAACCACCTGGACGCCCAG GCCGGCCTCACA
JCV241	GATGCCCATACCGCTCTGCGGCATGAACCCGATCGCGTGCACCACACCGTCGATCTTGT TGCCCTCACCGA
JCV242	GATGCCCATACCGCTCTGCGGCATGAACCCGATGGAGTGCACCACACCGTCGATCTTGT TGCCCTCACCGA

JCV243	GCTTTTCTGCGTTCTCGGGTAGCCGCT
JCV244	CAACAGCGCTAGCATCCTTGAGAGTT
JCV245	CAGGTTCGGTGGCGCGCTACGAATCT
JCV246	CGGCCAGCACGATGCGCGCGGATGCGT
JCV247	CCGCTCTGCGGCATGAACCCGATCGCGTGCACCACACCGTCGATCTTGTTG
JCV248	GCATGAACCCGATCGCGTGCACCACACCGTC
JCV249	CGCACGCTTAAGCCGACACGGTCAT
JCV250	CACGGTTTCTAGAACCGGCCACCGGA
JCV251	CGCGTCGAAGCTTCAGCTGCGGGCCCT
JCV252	GTGCGCGCAAGATCTCGCCCAGCAGCA
JCV253	CATGGACCAGAACAACCCGCTGTCGGGTCTGACCCGCAAGCGTCGTCTTTCGGCGCTGG GCCCCGGCGGTC
JCV254	GACCGCCGGGGCCCAGCGCCGAAAGACGACGCTTGCGGGTCAGACCCGACAGCGGGTT GTTCTGGTCCATG
JCV255	CACTACGGACCGCTGTTCATCCGCATGGCCTGGCAGGCCGCGGGCACCTACCGCGTCA GTGACGGCCGCGG
JCV256	CCGCGGCCGTCACTGACGCGGTAGGTGCCCGCGGCCTGCCAGGCCATGCGGATGAACA GCGGTCCGTAGTG
JCV257	CACTACGGACCGCTGTTCATCCGCATGGCCTGGTAGGCCGCGGGCACCTACCGCGTCAG TGACGGCCGCGG
JCV258	CCGCGGCCGTCACTGACGCGGTAGGTGCCCGCGGCCTACCAGGCCATGCGGATGAACA GCGGTCCGTAGTG
JCV259	CGAGACGATGGGTAACTACCATCCGCACGGCGACGTCTCGATCTACGACACCCTGGTC CGCATGGCCCAGC
JCV260	GCTGGGCCATGCGGACCAGGGTGTCGTAGATCGAGACGTCGCCGTGCGGATGGTAGTT ACCCATCGTCTCG
JCV261	GTGATCGGCGCCAACTCGTCCGACGACGGCTACATGCTGCAGATGGCGCGCACGGCCG AGCACGCGGGCTA
JCV262	TAGCCCGCGTGCTCGGCCGTGCGCGCCATCTGCAGCATGTAGCCGTCGTCGGACGAGTT GGCGCCGATCAC
JCV241	GATGCCCATACCGCTCTGCGGCATGAACCCGATCGCGTGCACCACACCGTCGATCTTGT TGCCCTCACCGA
JCV263	CGAACAAACCGTCCTCGAAACCCGTTCGGGTCACCTACTACAGGTAGACCGCCGGGAT GATCTGGTCGGTGTCGA
JCV264	TCGGGGCGGGCAACAAGCTCGACGGGGTGGTGCATGCGATTGGGTTCATGCCGCAGAC CGGGATGGGCATC
JCV265	GATGCCCATCCCGGTCTGCGGCATGAACCCAATCGCATGCACCACCCCGTCGAGCTTGT TGCCCGCCCCGA
JCV266	GTCGAAGAACGGGTTGATGCCCATCCCGGTCTGCGGCATGAACCCAATCGCATGCACC ACCCCGTCGAGCTTGTTGCCCGCCCCGATCGCCTCGGTCACCC
JCV267	CACTACGGGCCGCTGTTTATCCGGATGGCGTGGCAGGCTGCCGGCACCTACCGCATCCA CGACGGCCGCGG
JCV268	GTATGGCACCGGAACCGGTAAGGACGCGATCACCACCGGCATCGAGGTCGTATGGACG AACACCCCGACGA
JCV269	GCCACTACGGGCCGCTGTTTATCCGGATGGCGTGGTAGTAGGCCGGCACCTACCGCATC CACGACGGCCGCGGCGG
JCV270	GTGCCCGAGCAACACCCACCCATTACAGAAACCACCACCGGAGCCGCTAGCAACGGCT GTCCCGTCGTGGGAAGTTCGTGCAGGACTTCGTCGCTGCCTGGGACAAGGTGATGAAC CTCGACAGGTTCGACGTGCGCTGA
JCV271	GACCACCTCGCAGCCGTGGTGGCCCGCCGACTACGGCCACTACGGGCCGCTGTTTATCC GGATGGCGTGGGCACCTACCGCATCCACGACGGCCGCGGCGGCGCCGGGGGCGGCATG CAGCGGTTCGCGCCGCTTAACAG
JCV272	CGAACAAGCCGTCCTCGAAACCGGTTCGGGTGACCTACTACAGAAAGACCGCGGGAAT GATCTGATCGGTGTCGA
JCV273	ATCACGTCGAGCGGCGGCGGCGCCGCAGCGGCGGCCTACTAGATCTCCGACAGGCTCA CCGAGGCTTCACGCGCGG
JCV274	TTATTTTTGACACCAGACCAACTGGTAATGGTAGCGACCGGCGCTCAGCTGGAATTCCG CCGATACTGACGTTTGAGGGGACGACGACAGTATCGGCCTCAGGAAGATCGCACTCCA GCCAGCTTTCCGGCACCGCTTCT
JCV275	GCATCTGCCAGTTTGAGGGGACGACGACAGTATCGGATCGCACTCCAGCCAGCTTTCCG GCACCGCTTCT
JCV276	GGATAGGTCACGTTGGTGTAGATGGGCGCATCGTAGATCGCACTCCAGCCAGCTTTCCG

	GCACCGCTTCT
JCV277	TGCATCTGCCAGTTTGAGGGGACGACGACAGTATCCTACTAAGGAAGATCGCACTCCA GCCAGCTTTCCGGCACCG
JCV278	GATCCGCACCGTCGAGCAGTCCGACA
JCV279	GGCTCGACTACCGTTTCGGATTGCT
JCV280	GCCATCAATGAAAGAGCAACTGGCA
JCV281	CCAGCTTTACTCAGGCTGCGCACCA
JCV282	GGTGATGCCAGCGATGCGCAGTTCA
JCV283	CAGGAATCCAAGAGCTTTTACTGCTT
JCV284	GAAAGCTTCGAATTCTGCAGCTGGATC
JCV285	CCAAGACAATTGCGGATCCCGTCGT
JCV286	CGGCGATCCGGTCGTCGACGGGAGCGGCGGAAGCCTACTACATACGCACACCGGCGGC CGCCATCACTGCCAGGG
JCV287	TGTCGCGGTGCAGCAGCACCAGCGCGTCTTCGAGATCTGCTTCCTGCGAGTCGAAGTCG GTGACGAAGTAG
JCV288	GACAATTGCGGAGCTAGCCATGGACAT
JCV289	GTCATACGCGGCTAGCGGATCCCGTTA
JCV290	GCCTTCGGATCCTCCCCTGACGTGTA
JCV291	CGTTGTAAGCTTCGGGTGGATGTCA
JCV292	GCCATCATGGCCGCGGTGATCAGCTA
JCV293	GGTACCGAGCACTAGTTGACATAAGC
JCV294	TGTGCGTATGCCGAC
JCV295	TGTGCGTATGTAGTA
JCV296	CAGTGCACGCCGAGTTCGGGCAGCA
JCV297	GGTCTACCTGAAGCG
JCV298	GGTCTACCTGTAGTA
JCV299	GTGCGCCAGAGATAACCGCCTTGAACT
JCV300	CGACGGTGTGGTGCACT
JCV301	CGACGGTGTGGTGCAGG
JCV302	CGTAGATCACGGTGCCGGTGGT
JCV303	GATGCCCATACCGCTCTGCGGCATGAACCCGATCGCATGGACCACACCGTCGATCTTGT TGCCCTCACCGA
JCV304	CGACGGTGTGGTCCATG
JCV305	GACTCGCAGGAAGCAGT
JCV306	GACTCGCAGGAAGCACA
JCV307	CACGATGCCCTCCTCGACCGCT
JCV308	GACTCGCAGGAAGCAGATCTCGAAGAC
JCV309	GCCGGTGTGCGTATGTAGTAGGCTTCCGCCG
JCV310	CCGGCGGTCTACCTGTAGTAGGTGACCCGAA
JCV311	CCGGCGGTCTACCTGAAGCG
JCV312	CCGGCGGTCTACCTGTAGTA
JCV313	GCCGGTGTGCGTATGCCGAC
JCV314	GCCGGTGTGCGTATGTAGTA
JCV315	GTCCTCCCTATCAGTGATAGATA
JCV316	CTACTTCGTCACCGACTTCGACTCGCAGGAAGCAGATCTCGAAGACGCGCTGGTGCTGC TGCACCGCGACA
JCV317	TACCTCGAGGTCACCGAGGGCGTCGGGTTCGACAAGGGCTTCCTGTCGGCCTACTTCGT CACCGACTTCGACTCG
JCV318	CTCGGCGACCTTCTCCAGCAGCGGCAGCAGATCCGGCAGCGAGCTGATCTTGTCGCGGT GCAGCAGCACCAGCGC
JCV319	GCTGGAACTCGACGTGCAGAACGAGGAGCACCTGTCGACTCTGGCCGACCGGATCACC GCCGAGATCGGTGAGGG
JCV320	TACGAGTACGCCGAGATGTGGATGCCCTTGGACACATCCTCGTACGGCGCGTCGAAGA ACGGGTTGATGCCCATA
JCV321	GACGAATCTCTCACGACGCAGTGT
JCV322	TCAGCGCACGTCGAACCTGTCGAGGTTCATCACCTTGTCCCAGGCAGCGACGAAGTCCT GCACGAACTTCCCACGACGGGACAGCCGTTGCTAGCGGCTCCGGTGGTGGTTTCTGTAA TGGGTGGGTGTTGCTCGGGCAC
JCV323	CTGTTAAGCGGCGCGAACCGCTGCATGCCGCCCCCGGCGCCGCCGCGGCCGTCGTGGA TGCGGTAGGTGCCCACGCCATCCGGATAAACAGCGGCCCGTAGTGGCCGTAGTCGGCG GGCCACCACGGCTGCGAGGTGGTC
JCV324	CCGCCGCGGCCGTCGTGGATGCGGTAGGTGCCGGCCTACTACCACGCCATCCGGATAA

	ACAGCGGCCCGTAGTGGC
JCV325	CATGGACCAGAACAACCCGCTGTCGGGGTTGACCCGCAAGCGCCGACTGTCGGCGCTG GGGCCCGGCGGTC
JCV326	GACCGCCGGGCCCCAGCGCCGACAGTCGGCGCTTGCGGGTCAACCCCGACAGCGGGTT GTTCTGGTCCATG
JCV327	CCCGCTGTCGGGGTTGACCCACAAGCGCCGACTGTTGGCGCTGGGGCCCGGCGGTCTGT CACGTGAGCGTG
JCV328	CACGCTCACGTGACAGACCGCCGGGCCCCAGCGCCAACAGTCGGCGCTTGTGGGTCAA CCCCGACAGCGGG
JCV329	TGGTGTATGCACCCGCGTGTACACCACCACTCCGAGGAAGCCGAACTCGGCGCTTCGG AAGGTTGCCCGCG
JCV330	CGCGGGCAACCTTCCGAAGCGCCGAGTTCGGCTTCCTCGGAGTGGTGGTGTACACGCG GGTGCATACACCA
JCV331	GCCATCCTGACGGATGGCCT
JCV332	GCATGCGGATCGTGCTCATT
JCV333	CGCCGCCCGAAATGAGCACGATCCGCATGCCACCGCACCCATCAGAGATGGT
JCV334	TAAAAAAGGGGACCTCTAGGGTCCCCAATTAATTAGTTGTTCCTTTCGGGTGGATGTCA
JCV335	CGCCGCCCGAAATGAGCACGATCCGCATGCGCACACACCCCTGACTCCTGCTA
JCV336	CGAGCCTGCGGAACGACTAGGAATTCTGGGAGCCG
JCV337	CGGCTCCCAGAATTCCTAGTCGTTCCGCAGGCTCG
JCV338	GAAGGTGACCAAACCATATGACCACCA
JCV339	CATGCTGGAATTCGGGGCGATCATT
JCV340	TGCTGACATGCGGGCGTAGCTCAATGGTAGAGCCCTAGTCTTCCAAACTAGCTACGCGG GTTCGATTCCCGTCGCCCGCTCGGTAGGGACCGCCACGTGCGATTTAGGATACATGCTA GCCACCT
JCV341	GTGCGCCGATTTCTGCACCACGGTCGTGATCTGCGACGAACCACGACCTTGGTGCAGAA ATCGCGGGGGCAGTTGAGCACTCGGCAACGAAAAAGGGACCACCGCAGCGCTAGCGA GAACGTCCC
JCV342	ACCAGATCATCCCGGCGGTC
JCV343	ACCAGATCATCCCGGCGGGG
JCV344	CAGAAGGCCATCCTGACGGATGGCCT
JCV345	GCATGCGGATCGTGCTCATTTCGG
JCV346	GATTAGCTAAGCAGAAGGCCATCCT
JCV347	CTACGCGAGCCTGCGGAACGACTAGGAATTCTGGGAGCCGCTGGC
JCV348	GCCAGCGGCTCCCAGAATTCCTAGTCGTTCCGCAGGCTCGCGTAG
JCV349	GCTCGACGGGGTGGTGCAAT
JCV350	GCTCGACGGGGTGGTGCAAG
JCV351	CAGACAGCAGCGCGCACACCGTCTT
JCV352	GCTCGACGGGGTGGTGCATGCGATTGGGT
JCV353	CCCGCGGTCTTTCTGAAGCG
JCV354	CCCGCGGTCTTTCTGTAGTA
JCV355	GAGTTTCCGCAGCGTAAGGGCTAT
JCV356	CCCGCGGTCTTTCTGTAGTAGGTCAC
JCV357	GATCCAATATTACTAGTAGATCTCGT
JCV358	GACGTCTTAATTAATATGCATCAAT
JCV359	GACGTCTTAATTAATATGCATCAATTGATTTA
JCV360	GATCCAATATTACTAGTAGATCTCGTAATATTG
JCV361	GAATAGAGGTCCGCTGTGACATAGGAATCCCTGTTACTTCTC
JCV362	GAGAAGTAACAGGGATTCCTATGTCACAGCGGACCTCTATTC
JCV363	GACCTCTAGGGTCCCCAGCTGGCTAG
JCV364	CACGGCCGTGACGCTAGCGACGATCCA
JCV365	AAAACGATTGTCATTATCGTACGACGGTACCGCACGACGAAGGAGAGTCAATGGCTCG CAACGAGATCCGGCCCATCGTGAAGCTGCGGTCCACTGCGGG
JCV366	CCCGCAGTGGACCGCAGCTTCACGATGGGCCGGATCTCGTTGCGAGCCATTGACTCTCC TTCGTCGTGCGGTACCGTCGTACGATAATGACAATCGTTTT
JCV367	AGAAGGTCTGATGGCTCGCAACGAGATCCGGCCCATCGTGAAGCTGCGGTATGGCGAA GAAGTCGAAGATTGTCAAGAACGAGCAGCGGCGAGAACTGGT
JCV368	ACCAGTTCTCGCCGCTGCTCGTTCTTGACAATCTTCGACTTCTTCGCCATACCGCAGCTT CACGATGGGCCGGATCTCGTTGCGAGCCATCAGACCTTCT
JCV369	AAATACGATCCGGTCCTGCGCCGCCACGTCGAGTTCCGCGAGGAACGCTGATGGCAGT CAAGAAGTCCAGAAAGCGCACGGCCGCAACTGAACTCAAGAA
JCV370	TTCTTGAGTTCAGTTGCGGCCGTGCGCTTTCTGGACTTCTTGACTGCCATCAGCGTTCCT

	CGCGGAACTCGACGTGGCGGCGCAGGACCGGATCGTATTT
JCV371	GAGATGGCGCATCGCGGCGAGTTGCCCGGTGTGCGGAAGGCGAGTTGGTGGGCGTGCG GTATGACACCATCGGTGCCGAAGGCGACTGCGGATCGAGGAA
JCV372	TTCCTCGATCCGCAGTCGCCTTCGGCACCGATGGTGTCATACCGCACGCCCACCAACTC GCCTTCCGCACACCGGGCAACTCGCCGCGATGCGCCATCTC
JCV373	GTGGCTGACGAGCAGGGTGCCAGGT
JCV374	CACCGATGGTGTCATACCGCACGCCT
JCV375	ATCATCCCGGCGGGGCCTTACGCT
JCV376	GGAAAACCCTGGCGTTACCTAGCTTAATCGCCTTGCAG
JCV377	CTGCAAGGCGATTAAGCTAGGTAACGCCAGGGTTTTCC
JCV378	GTCTCTGACGAGCGGGAGAACCCA
JCV379	GAGCTTCAACCCACCATATGAGGAAGGCA
JCV380	GTCACGACCGGTTGTGTGAGCCAGA
JCV381	CTGTCCTCGTTGGGTACCGAGCTCGA
JCV382	CGACGAAGGAGAGTCAATCT
JCV383	CGACGAAGGAGAGTCAATCG
JCV384	GATCCGGAGGAATCACTTCGCAATG
JCV385	TGATGTCGCTGCAGGAACTGCACAGCGAACTGGGGTCGCGCCGGTCATGACGGGCCCA CCACGCGACAGCGCCATTGCGCCGACAATCGTCGGTCCAACG
JCV386	CGTTGGACCGACGATTGTCGGCGCAATGGCGCTGTCGCGTGGTGGGCCCGTCATGACC GGCGCGACCCCAGTTCGCTGTGCAGTTCCTGCAGCGACATCA
JCV387	CTCGGCCGCGGTGCAGGGCATCGAGGCCGGCATCCGCGGCGACATCGGCGTGATGTCG CTGCAGGAACTGCACAG
JCV388	GCACAGGGCTGCCGGCTCGGACACATCTGAAATCCGGCTCCTACCTGAGCCGTTGGAC CGACGATTGTCGGCGCA
JCV389	CATCCCGTGCGTCACCACGGTGCA
JCV390	GTGATGGACTGCGCCGCAGGCGAACT
JCV391	GGCGACCAGCACTACCGGCGTACGCATGGGACCTCCCGGTTTGCTTATTGAAAACGATT GTCATTATCGTACGAC
JCV392	GGTCCGGGTCGTTGCGGCGATTCTTGCGGGTGACGTACGTGTATCCGGTTCCCGCAGTG GACCGCAGCTTCACGA
JCV393	CCGTAGGCCTCGAAGAACGCGACCTACGGTTTGACGATGGCGAAGCCCGCG
JCV394	CCGTATGACTCAAAAAACGCCACCTACGGTTTGACCACCGCGAAATCAGCG
JCV395	CCCCGCCGAACCGTAGGCCTCGAAGAACGCGACCTACGGTTTGACGATGGCGAAGCCC GCGAACGCCGCGA
JCV396	GGCGTCGACCGTCAGGCCCCAGGCGTTCAAGAGCTACTAATGCGGGTCGATGCCGGGG CACAGCGGCCCGCGCG
JCV397	TCCTCCCACTCCCTCGGTCTCACCT
JCV398	TCGGCCGCGGTGCAGGGCATCGAGG
JCV399	CGAAGAGGCCGACCCGATTGAAGGGGATTACATCTATGGCTGAAAATGCTGGGCCCAA CGCATGAGCGCCCCGGCGAACCACGACGCGGTGGTTGATCTG
JCV400	ACCGTCGGCGACGTCGTCACCGACAGCTATATCTACGACACCGACCCGCTCGAAGAGG CCGACCCGATTGAAGGG
JCV401	GGCCTCGGCGTAGTGGTACGTCGGCGGGCGCGTCATGACTGCACCACCTGCAGATCAA CCACCGCGTCGTGGTTC
JCV402	GACGTGGTCACGATCAGCCTGCCCT
JCV403	CAACTTGAACGCGATCGCGGGCACG
JCV404	GCTCCGGGCTCGCAGCAGCGGGCTT
JCV405	CAACGCAATTAATGTGAGTTAGCTCA
JCV406	GGGTGATGTCCGGGGCGGTACGCGTAGGCGTCTGAAAGAGACTTATGAGCGTTCTCAA CGCCGCCTCGGTGGCCCGGATGATCCTGACGACGGAGACCGC
JCV407	GCGGTCTCCGTCGTCAGGATCATCCGGGCCACCGAGGCGGCGTTGAGAACGCTCATAA GTCTCTTTCAGACGCCTACGCGTACCGCCCCGGACATCACCC
JCV408	GCGCAGTCGATCGTCGACGCGGTCGCGCAGGCGAACCGGGAGGCGGATCCGGCGGCGC GCGACGGCGATCCGGTGGGCCCGTTCGGTGTCGTGGTGGGCG
JCV409	CGCCCACCACGACACCGAACGGGCCCACCGGATCGCCGTCGCGCGCCGCCGGATCCGC CTCCCGGTTCGCCTGCGCGACCGCGTCGACGATCGACTGCGC
JCV410	GAGGAGTCAACCCCATATGATGTCGAT
JCV411	GGTCGGCGAATTCCGTTGTCATGTC
JCV412	AGCTGCACATATGCGCTGCGCCCGT
JCV413	GGATAGGTCAGGAATTCGCCGGTCA
JCV414	GTCGCGCCATATGGTGCAGATTCT

JCV415	GGTGGCGCTGAATTCGTCGTACGTCA
JCV416	GAGGTGGCGTCCATATGCACATGCA
JCV417	CGTCTGCTGCTTGAATTCGGTCACAG
JCV418	CGGATGCTCACGGATCCGATCTGCCA
JCV419	GATCTTGGCGATGAATTCGCTGCGCT
JCV420	GAGGGCGGCATATGAGCGCCGA
JCV421	CTGTCCCTATCGAATTCGGCCTGGT
JCV422	GGCGCATATGAGCCGCGAGCTGCT
JCV423	CTTCGCCTCGTCGAATTCCGGGTCGT
JCV424	GTGCAATGACATATGAGTGACGGAC
JCV425	CATCGTTTGTGAATTCTCGCTTGAA
JCV426	GGATCGACATATGAAGCGCACCAGGA
JCV427	CTCGCCGGAATTCTGCTACGGGTCGA
JCV428	GCGGTCGACATATGCTGAGCGTGCAGCC
JCV429	CTGCATAACCGGAATTCAGGGCGTGT

6.5 PROTEIN PURIFICATION

C-terminally 6x-his-tagged Che9c gp60, Che9c gp61, and Halo gp43 proteins were purified by over-expression from T7-inducible vectors in *E. coli* BL21(DE3) pLysS (Invitrogen) transformed with pJV33 and pJV34, respectively, using nickel-affinity chromatography. In addition, empty vector pET21a was also transformed into BL21(DE3) pLysS cells and prepared in parallel as a negative control.

Che9c gp60 was purified as follows: cells containing pJV33 were grown at 37°C to an OD_{600} of 0.4. Che9c gp60 expression was induced by addition of 1 mM IPTG, and cells were harvested by centrifugation after 4 hr growth at 30°C. Cell pellets were resuspended in lysis buffer at 5 ml per gram wet weight and sonicated 6-8x at bursts of 10 s with 1 min cooling in between bursts. Cell lysate was centrifuged at 13,000 rpm for 30 min at 4°C, and the supernatant was incubated with 750 µl Ni-NTA agarose resin (Qiagen; washed 2x with dH_2O) rocking at 4°C for 1.5 h. This was applied to a column and washed 2x with 20x column volumes of wash buffer containing 10 mM imidazole (first wash) or 20 mM imidazole (second wash). Protein was eluted

4x with 1 column volume elution buffer. Eluted fractions were analyzed by SDS-PAGE, and the fractions containing the least contaminating proteins were dialyzed against storage buffer.

Che9c gp61 purification was performed similarly except 2.5 ml Ni-NTA agarose resin was used, washes did not contain imidazole, and elutions were performed with a gradient of imidazole concentrations (20, 40, 60, 80, 100, 150, 200, and 250 mM imidazole). Elution fractions with 150 and 200 mM imidazole were dialyzed against storage buffer. Proteins were analyzed by SDS-PAGE and quantified by Bradford protein assay (Bio-rad).

Halo gp43 protein purification was performed similarly to that for gp61 except 5 ml of resin was used. Two combinations of elution fractions (80 and 100; 150 and 200 mM imidazole) were dialyzed against storage buffer and analyzed and quantified as above.

6.5.1.1 Antibody synthesis

Anti-gp61 and anti-gp43 antibodies were synthesized by Pocono Rabbit Farms. Screening bleeds were initially tested from rabbits and found to have very high background on wild type *M. smegmatis* cell extracts. Therefore, multiple mouse screening bleeds were tested, and three mice were chosen for each protein (Che9c gp61 and Halo gp43). Approximately 1 – 1.5 mg of nickel-affinity purified protein was sent to Pocono Rabbit Farms, and mouse bleeds were obtained. Mice S12 and S14 were inoculated with Che9c gp61 protein (mouse S13 died), and mice S19, S20, and S21 were inoculated with Halo gp43 protein. Titermax adjuvant was used in place of Freund's, which is not suitable for mycobacteria. Approximately four bleeds were collected for each mouse, and subsequently the final bleeds were extracted. All antibodies were stored in 5 μl aliquots at -80°C.

6.6 IN VITRO ASSAYS

6.6.1 Exonuclease assays

Exonuclease activity of Che9c gp60 was determined as described [227] by three assays - A, B, and C - in which λ Exo (5 U/μl; NEB) was used as a positive control for DNA degradation, while mock purified pET21a vector control protein preparations were used as a negative control. Assays A and B were used to determine the exonuclease activity of Che9c gp60 on linear dsDNA, while assay C was used to determine its activity on supercoiled or nicked open-circle dsDNA.

In assay A, reaction mixtures (50 μl) contained 8 nM ^{32}P-labeled 100 bp Bxb1 *attB*, 40 nM unlabeled 100 bp Bxb1 *attB* (see Table 16), and gp60 or control proteins in exonuclease assay buffer. Che9c gp60 (10.6 pmol/μl), λ Exo, and mock purified control protein were added at 1 μl per reaction of serially diluted concentrations (neat, 2x, 4x, 8x, 16x, and 32x in protein dilution buffer) with respect to the stock concentration of each protein. Reactions were incubated for 5 min at room temperature and stopped with final concentrations of 10 mM EDTA and 0.5% SDS and addition of glycerol loading dye (0.25% bromophenol blue, 0.25% xylene cyanol, and 30% glycerol). Samples were analyzed by gel electrophoresis on 8% native TBE polyacrylamide gels, dried on Whatman paper for 1.5 hr at 85°C, and exposed to BioMax Maximum Resolution film (Kodak) overnight.

In assay B, reaction mixtures (50 μl) contained 0.8 nM pBluescript SK$^+$ (Stratagene) linear dsDNA (linearized by digesting with *Eco*RV) and protein in exonuclease assay buffer. Che9c gp60 and mock-purified protein were added at equal volumes (1 μl) from parallel

purifications, with gp60 at a final concentration of 212 nM; λ Exo was added at 5 U per reaction. Negative control reactions were routinely performed using 1 μl storage buffer instead of enzyme. Reactions were incubated for 0 – 10 min at room temperature, stopped with EDTA and SDS as above, and loading dye was added. 25 μl of each reaction was analyzed by gel electrophoresis in 0.8% agarose gels, visualized with ethidium bromide (0.4 μg/ml) and photographed using a Bio-Rad Gel Doc 2000 and Quantity One 4.6 software.

In assay C, reaction mixtures (20 μl) contained either 2 nM linearized or 2 nM circularized pBluescript SK^+ and protein in exonuclease assay buffer. Che9c gp60 was added to a final concentration of 530 nM by addition of 1 μl protein and 1 μl mock-purified control protein was used; λ Exo was added at 5 U per reaction. Reactions were incubated for 0 – 10 min at room temperature, stopped by addition of EDTA and SDS as above, and loading dye added. Reactions were analyzed on agarose gels as with assay B.

6.6.2 DNA binding assays

6.6.2.1 Double-filter binding assay

Filter-binding assays were used to determine the DNA binding activity of Che9c gp61 and Halo gp43 [227]. A double-filter binding assay was utilized [20,239] in which nitrocellulose filter membranes were stacked on top of DEAE-cellulose filters and filtered by means of a FH 225 V filter manifold system (GE Healthcare). Protein and protein/DNA complexes are bound by the nitrocellulose filter on the top, and the unbound DNA is captured by the DEAE filter below, so that the total DNA counts per sample are accounted for. DEAE-cellulose filters, grade DE81 (Whatman), were prepared as described [20] in which the membranes were treated with 0.1 M

Na_2EDTA pH 8.8 for 10 min rocking at room temperature, followed by 3-10 min washes in 500 mM NaCl. The filters are then briefly treated with 500 mM NaOH (less than 30 sec), washed 15 times with dH_2O, washed with TBS buffer, and stored at 4°C in that same buffer. Protran BA85 nitrocellulose filters (Whatman/Schleicher & Schuell) were prepared by removing the blue packaging papers, and incubating the filters in 500 mM KOH on a rocking platform for 25 min (time critical) at room temperature. The filters were washed 15 times with dH_2O, washed twice with TBS buffer, and stored at 4°C in TBS buffer.

DNA binding reactions (30 µl) contained ^{32}P-labeled Bxb1 ss-*attB* (50 nt or 100 nt; Table 16), enzyme, and binding assay buffer similar to the conditions described for assaying DNA binding of RecT [68]. Reactions containing dsDNA (~66.7 nM) used an annealed ^{32}P-labeled 50 bp *attB* substrate (Table 16) which was gel-extracted from a native 8% TBE polyacrylamide gel (exposed to film for 30 sec to locate probe), electroeluted in dialysis membrane (SpectraPor, MWCO 1000) in TBE at 100 v for 45 min, phenol:chloroform extracted and ethanol precipitated. In most reactions, 90% of the DNA added was cold, and 10% was used 'hot' to reduce the amount of radioactivity use. The concentrations of gp61 used were in the range of 0.2 – 3.7 µM (diluted in protein dilution buffer), and storage buffer was used as a 'no protein' control. Reactions were incubated at 37°C for 20 mins and filtered. Each reaction (25 µl) was filtered and washed once with wash buffer. Filters were individually placed in 5 ml ScintSafe 30% scintillation fluid (Fisher) and the radioactive counts per minute (CPM) were determined by counting in a scintillation counter, with 1 min counting per sample. Percent DNA bound was determined by the following equation:

$$\%DNA_{bound} = \frac{CPM_{NC}}{(CPM_{NC} - CPM_{DEAE})}$$

C_{NC} is the CPM retained on the nitrocellulose filters (protein and protein/DNA complexes), and CPM_{DEAE} is the CPM on DEAE filters (unbound DNA), and this is further normalized to negative controls in which no protein was added. Binding constants were determined by non-linear regression using SigmaPlot 8.0. Error bars represent the standard deviation from three independent experiments.

6.6.2.2 Gel shift assay

Gel shift assays were performed with the DNA binding reaction conditions similar to those described for the filter-binding assays using either ss-*attB* (50 nt) or ds-*attB* substrates (Table 16). DNA binding reactions were loaded on native 8% TBE polyacrylamide gels with glycerol loading dye, dried on Whatman paper for 1.5 hr at 85°C, and exposed to BioMax Maximum Resolution film (Kodak) overnight or on a Fujifilm phosphoimager plate. Phosphoimager plates were scanned with a Fujifilm FLA-5100, and quantitation of shifted bands was performed using MultiGauge software.

6.6.3 Electron Microscopy

Samples were prepared and analyzed by transmission electron microscopy as described [20] with reactions (50 µl) containing purified Che9c gp61 [227] at a concentration of 50 µg/ml (1.2 µM) and 1.95 µM ss-*attB* DNA (100 nt; Table 16) or oligonucleotides of various lengths. Briefly, the reactions were absorbed to glow-discharged, formvar carbon-coated 400 mesh copper grids, and these were negatively stained with 2% uranyl acetate. Images were collected at a magnification of 140,000x.

6.6.4 Gel filtration

Analytical gel filtration of purified Che9c gp61 protein was performed using a Superdex 200 10/300 GL high performance gel filtration column (Tricorn, Amersham) controlled by the System Gold high-pressure liquid chromatography system (Beckman Coulter, Inc.) as previously described [101]. The column was standardized using a gel filtration calibration kit (Amersham) in which both low molecular and high molecular weight protein standards were run in duplicate to determine their respective elution volumes. Standards included: ribonuclease A (13.7 kDa), chymotrypsinogen (25 kDa), ovalbumin (43 kDa), albumin (67 kDa), aldolase (158 kDa), catalase (232 kDa), and apoferritin (443 kDa). Tryptophan and Barnyard Phage particles were used to determine the bed and void volumes, respectively. Approximately 330 μg of each standard was used for calibration at 100 μl volumes. The column was first equilibrated in one column volume (approximately 25 ml) of equilibration buffer (33 mM Tris, pH 7.5, 100 mM NaCl) at a flow-rate of 0.2 ml/min. Protein standards or gp61 protein were loaded through a 100 μl loop and eluted at a flow-rate of 0.5 ml/min. Elution volumes of proteins (recorded as the time of peak absorbance) were monitored at a wavelength of 280 nm using protein fluorescence and, as an additional reading, on a Jasco fluorescence detector. The K_{av} values (fractional retentions of the samples [68]) for each protein standard were calculated based on the following equation:

$$K_{av} = \frac{Ve - Vo}{Vt - Vo}$$

Ve is the elution volume of the protein, Vo is the column void volume, and Vt is the total bed volume of the column. The K_{av} value for each protein standard was plotted against the molecular

weight (Daltons) on a logarithmic scale, and a trendline was determined based on the standards. To assay the size of gp61 protein complexes, gp61 protein was tested at concentrations of 5 μM, 10 μM, and 25 μM. The elution volumes were collected for each, and the K_{av} values were determined. Using the equation from the slope of the line, and solving for y, the molecular weight for each gp61 reaction was determined. The column error was determined for the molecular weight of the protein standards compared to the value obtained using the slope equation in order to determine the range of accuracy for the gp61 reactions.

6.7 WESTERN BLOT ANALYSIS

Che9c gp61 and Halo gp43 protein expression were monitored using western blot analysis [227]. *M. smegmatis* cells were collected, resuspended in SDS-PAGE buffer and normalized to OD_{600}. To do this, the OD_{600} was measured (OD per ml), the cells pelleted, and resuspended to a concentration of 0.05 ODs per μl. Strain samples (0.5 ODs of cells) and purified protein (0.5 μg) as positive controls were loaded on 10% SDS-polyacrylamide gels following boiling for 3 – 5 min at 95°C in 4x SDS-PAGE loading dye diluted down to 1-2x. Proteins were transferred to Sequi-blot PVDF membranes (Bio-rad) using a Bio-rad semi-dry transfer cell in transfer buffer (48 mM Tris, 39 mM glycine, 0.037% SDS, 20% methanol) at 0.8 mAmps per cm^2 of gel. The blots were allowed to dry, rewetted with methanol, and blocked in 5% milk in TBS-T for 1 hr rocking at room temperature. These were probed with either mouse anti-gp61 polyclonal antibodies or mouse anti-gp43 polyclonal antibodies in 1% milk in TBS-T (1:5,000) and rocked for 1 hr at room temperature. The blots were then washed 3 times for 5 min in TBS-T, and

subsequently probed with sheep anti-mouse horseradish peroxidase-linked secondary antibody in 0.5% milk in TBS-T (1:15,000) for 45 min rocking at room temperature. The blots were washed again and submerged in 14 ml Western Lightening Reagent (Perkin-Elmer) for 2 mins to detect the secondary antibody. Blots were visualized using a Fujifilm LAS-3000.

6.8 SOUTHERN BLOT ANALYSIS

6.8.1 Genomic DNA preparation from Mycobacterial cultures

DNA was isolated from saturated *M. smegmatis* and *M. tuberculosis* cultures as described [229]. Briefly, 10 ml of culture was centrifuged at 2,000 x g for 20 min, and the cell pellet was resuspended in 1 ml GTE Solution (25 mM Tris-HCl pH 8.0, 10 mM EDTA, 50 mM glucose). This was transferred to a microcentrifuge tube and pelleted for 10 min, and resuspended in 450 µl GTE Solution. To this, 50 µl of lysozyme solution (10 mg/ml in Tris pH 8.5) was added, gently mixed, and incubated at 37°C overnight. 100 µl 10% SDS was then added and gently mixed, and 50 µl 10 mg/ml Proteinase K (Sigma) was added and mixed gently; this was incubated at 55°C for 20 to 40 min. 200 µl 5M NaCl was then added and gently mixed, and 160 µl of CTAB (preheated at 65°C) was added and mixed gently. This was incubated at 65°C for 10 min. Finally, an equal volume (~1 ml) chloroform:isoamyl alcohol (24:1) was added to the tube, the aqueous phase containing the DNA was extracted and precipitated similar to the procedure for preparing mycobacteriophage genomic DNA (see below), and this was stored at -20°C. When applicable, the BSL3 *M. tuberculosis* strains were only removed from the BSL3 lab following extraction with chloroform:isoamyl alcohol.

6.8.2 Southern blotting procedures

Southern blots were performed as described [7] to determine the genotype of a putative mutant strain, such as an allelic replacement mutant. The predicted sequence of the mutant strain was used to choose restriction sites that would result in different sized fragments (1 – 10 kbp) when the DNA was digested, making it possible to distinguish between wild type and mutant strains. Chromosomal DNA was isolated as described above and DNA digests (20 µl) were incubated 8 hr – O/N and run on an agarose gel to separate fragments.

The DNA was transferred to a nylon membrane using alkaline transfer conditions followed by downward capillary transfer. To prepare the gel, it was rinsed in dH_2O and placed in 0.25 M HCl on a rocking platform at room temperature for 30 min. The gel was rinsed with dH_2O and placed in 0.4 M NaOH on a rocking platform at room temperature for 20 min to denature the DNA. To setup for the downward capillary transfer, a stack of paper towels (~2 inches) topped with four pieces of Whatman paper (all cut to the size of the gel) were stacked next to a dish containing transfer buffer (0.4 M NaOH). The nylon membrane was placed on top of the Whatman paper (one piece wetted with transfer buffer), and the gel was gently placed on top of the nylon membrane, removing bubbles. Three wet pieces of Whatman were placed on top of the gel, and a long piece of Whatman was placed on top of this that extended down into the dish of transfer buffer. Two glass gel plates were placed on top of the stack to prevent evaporation and to provide some weight, and this was left to transfer for 2 hr or longer. After transfer, the wells were marked with a pencil, and the membrane air-dried. DNA was cross-linked to the nylon by using a Stratalinker UV Crosslinker (Stratagene) on auto-crosslink mode.

Membranes were treated with 20 ml of a pre-hybridization solution in a roller bottle at 50°C for 2 hr, and at this time the radioactive probe was prepared. A dsDNA substrate (typically 500 bp of PCR product) was used for hybridization to detect genotypic differences between strains tested by Southern blot. The DNA was denatured by boiling for 5 min at 95°C and cooled on ice, and a random primers labeling reaction (50 µl) was used to radiolabel the substrate. The reaction (ssDNA substrate, random primers mix, dTTP, dCTP, dGTP, [α-^{32}P]dATP, and Klenow enzyme) was incubated for 1 hr at 30°C, denatured at 95°C, and cooled on ice. Denatured probe and 1 mg denatured sheared salmon sperm DNA were added to 10 ml hybridization solution. The membrane was transferred to this solution and incubated in a roller bottle overnight at 65°C. The membrane was washed one time in wash solution 1 for 20 min at room temperature rolling, and two times in wash solution 2 for 20 min., first at room temperature and then at 65°C rolling. The membrane was exposed to film overnight and developed.

6.9 BACTERIAL STRAINS, GROWTH CONDITIONS, AND MANIPULATIONS

6.9.1 *Escherichia coli*

6.9.1.1 Strains/Media

E. coli DH5α, *E. coli* GC5, and *E. coli* 5α strains were grown in LB broth or on LB agar plates supplemented with kanamycin (Kan, 20 µg/ml), hygromycin (Hyg, 150 µg/ml), carbenicillin (Cb, 50 µg/ml), tetracycline (Tet, 6.5 µg/ml), chloramphenicol (15 µg/ml),

gentamicin (Gent, 15 μg/ml), and/or x-gal (40 μg/ml) as required. Strains were stored in 20% glycerol at -80°C.

6.9.1.2 Transformations

Transformations of all chemically competent *E. coli* cells were performed according to manufacturers' instructions for GC5 or DH5α cells. Briefly, DNA was added to 50 μl cell aliquots on ice for 30 min, heat-shocked at 42°C for 45 sec, recovered in 450 μl TSB for 1 hr, and plated on selective media. Electrocompetent cells (typically XL1-Blue) were transformed by thawing cell aliquots on ice, adding DNA to cells, and incubating on ice for 10 min. Cells were transferred to chilled 0.2 μM cuvettes (Bio-rad) and transformed with a Bio-Rad Gene Pulser II set at 2.5 kV, 200 Ω, and 25 μF. Cells were recovered in 1 ml TSB broth for 1 hr at 37°C, and plated on selective media.

6.9.2 *Mycobacterium smegmatis* mc^{2}155

6.9.2.1 Strains/Media

The high-efficiency transformation strain *M. smegmatis* mc^{2}155 was used for all manipulations [211]. *M. smegmatis* was grown in 7H9 broth (Difco) supplemented with 10% ADC, 0.2% glycerol, and 0.05% tween, and on 7H10 agar (Difco) plates supplemented with 10% ADC and 0.5% glycerol as described [18] unless otherwise mentioned. When required, media was supplemented with the following: Kan (20 μg/ml), Hyg (150 μg/ml), Cb (50 μg/ml), Chx (10 μg/ml), Tet (2.5 μg/ml), Gent (10 μg/ml plates; 2.5 μg/ml liquid), isoniazid (INH; 25 μg/ml), ethionamide (ETH; 25 μg/ml), rifampicin (Rif; 200 μg/ml), streptomycin (Str; 20 μg/ml),

ofloxacin (Ofx; 0.5 μg/ml), ethambutol (10 μg/ml), x-gal (40 μg/ml), 5-fluoroorotic acid (5-FOA; 1 mg/ml), uracil (0.2 mM), and/or leucine (100 μg/ml). Single colonies were picked and routinely inoculated from streak plates into 3 ml 7H9 broth with ADC, tween, and the appropriate antibiotic, and these were incubated with shaking at 37°C until saturated. Strains were stored at -80°C in 20% glycerol, and were streaked on 7H10 plates directly from these frozen glycerol stocks when required. *M. smegmatis* strains constructed are listed in Table 17.

Table 17. *M. smegmatis* strains.

Strain background	Relevant mutation(s)	Replicating plasmid	Antibiotic resistance	Recombineering substrate used to construct strain
M. smegmatis mc^2155	Δ*0642*:*res-hyg*^{R}*-res*	pJV24	Kan^R, Hyg^R	AES; pMP6
M. smegmatis mc^2155	Δ*4308*:*res-hyg*^{R}*-res*	pJV24	Kan^R, Hyg^R	AES; pMs4308
M. smegmatis mc^2155	Δ*6008*:*res-hyg*^{R}*-res*	pJV24	Kan^R, Hyg^R	AES; pPJM04
M. smegmatis mc^2155	Δ*groEL1*:*res-hyg*^{R}*-res*	pJV24	Kan^R, Hyg^R	AES; pMsgroEL1
M. smegmatis mc^2155	Δ*groEL1*:*res-sacB-hyg*^{R}*-res*	pJV53	Kan^R, Hyg^R	AES; pJV149
M. smegmatis mc^2155	Δ*leuB*:*res-sacB-hyg*^{R}*-res*	pJV53	Kan^R, Hyg^R	AES; p0004SleuB
M. smegmatis mc^2155	Δ*leuD*	pJV76amber	Kan^R, Hyg^R	100bp dsDNA
M. smegmatis mc^2155	Δ*leuD*:*res-sacB-hyg*^{R}*-res*	pJV24	Kan^R, Hyg^R	AES; p0004SleuD
M. smegmatis mc^2155	Δ*pyrF:gent*^{R}	pJV98	Kan^R, Gent^R	AES; pKP134
M. smegmatis mc^2155	Δ*recA*	pJV53	Kan^R	Unmarked with γδ res; removed pGH542
M. smegmatis mc^2155	Δ*recA*:*res-hyg*^{R}*-res*	pJV53	Kan^R, Hyg^R	AES; pJV28
M. smegmatis mc^2155	Δ*recB*	pJV53	Kan^R	Unmarked with γδ res; removed pGH542
M. smegmatis mc^2155	Δ*recB*:*res-hyg*^{R}*-res*	pJV53	Kan^R, Hyg^R	AES; pJV68
M. smegmatis mc^2155	Δ*recD*	pJV53	Kan^R	Unmarked with γδ res; removed pGH542
M. smegmatis mc^2155	Δ*recD*:*res-hyg*^{R}*-res*	pJV53	Kan^R, Hyg^R	AES; pJV101
M. smegmatis	*blaS* 25* 26*	pJV62	Kan^R	ssDNA (JCV286)

mc^2155				
M. smegmatis mc^2155	*gyrA* A91V	pJV62	Kan^R, OFX^R	ssDNA (JCV260)
M. smegmatis mc^2155	*inhA* S94A	pJV62	Kan^R, INH^R	ssDNA (JCV217)
M. smegmatis mc^2155	*rpoB* H442R	pJV62	Kan^R, RIF^R	ssDNA (JCV254)
M. smegmatis mc^2155	*rpsL* K43R	pJV62	Kan^R, Str^R	ssDNA (JCV219)

These are strains with specifically engineered mutations that were constructed by recombineering; the type of recombineering substrate used for each mutation is described. This list does not include strains constructed merely by introducing a replicating or integrating plasmid. Abbreviations: AES, allelic exchange substrate; res, resolvase site.

6.9.2.2 Competent cell preparations

Electrocompetent cells of *M. smegmatis* were made as described [18,227]. Briefly, cultures were grown to an OD_{600} = 0.8 – 1.0 and placed on ice for 30 min to 2 hr. These were centrifuged at 5,000 rpm for 10 min at 4°C, the supernatant was discarded, and the pellets were washed with ½ the original volume of ice-cold 10% glycerol. Centrifugation and washing of the cell pellets was repeated 2-3 times using 1/4, 1/8, and 1/10 volumes for washes. The final cell suspension was in 10% glycerol at approximately 1/15 – 1/25 the original volume, or between an OD_{600} = 5.5 – 7.0. After variation of experimental conditions, it seems that there is a window of cell concentration in which the highest level of competency can be achieved. Additionally, using larger wash volumes (*i.e.*, ½, ½, and ¼ in succession) and larger culture volumes (>50 ml) results in better cell competency. Cell aliquots were placed on dry ice and frozen at -80°C until use.

6.9.2.3 Transformations

Transformations of electrocompetent cells were performed by thawing competent cell aliquots on ice, using approximately 100 μl per transformation. DNA was added to cells, mixed

gently, allowed to incubate on ice for 10 min, and the cell mixture was transferred to chilled 0.2 μM cuvettes (Bio-rad). Cells were transformed with a Bio-Rad Gene Pulser II set at 2.5 kV, 1000 Ω, and 25 μF, typically with time constants above 20. Transformed cells were recovered for 2 hr or longer in 7H9 broth with ADC and tween shaking at 37°C. These were plated on 7H10 selective media, and incubated at 37°C for 3 – 5 days.

6.9.2.4 Assay for UV sensitivity

M. smegmatis strains to be tested for their phenotype following UV exposure were grown in the desired medium to an OD_{600} = 0.8. The assay was performed in two ways. In one approach, 1 ml of the culture was placed in a sterile Petri dish and exposed to UV at levels between 50 – 300 J/m^2 using the Stratalinker UV Crosslinker. The cells were subsequently serially diluted and plated on solid media. Alternatively, serial dilutions of the cultures were plated first and then subjected to UV treatment. Following either experiment, the plates were incubated at 37°C and colony numbers recorded wild type and Δ*recA* strains were always used as positive and negative controls, respectively.

6.9.3 *Mycobacterium tuberculosis*

6.9.3.1 Strains/Media

M. tuberculosis H37Rv and *M. tuberculosis* $mc^2$7000 were used for all manipulations. *M. tuberculosis* $mc^2$7000 is a derivative of H37Rv in which the RD1 region and *panCD* were both deleted, resulting in a pan- phenotype [150]. *M. tuberculosis* was grown in 7H9 broth (Difco) supplemented with 10% OADC (ADC plus oleic acid, BDL), 0.5% glycerol, and 0.05%

tween, and on 7H11 agar (Difco) plates supplemented with 10% OADC and 0.5% glycerol as described [18] unless otherwise mentioned. All experiments with *M. tuberculosis* mc^{2}7000 were performed with pantothenate added to media at 100 μg/ml. When required, media was supplemented with the following: Kan (20 μg/ml), Hyg (50 μg/ml), Cb (50 μg/ml), Chx (10μg/ml), INH (0.2 μg/ml), Eth (10 μg/ml), Rif (10 μg/ml), and/or Str (6 μg/ml). Strains were stored at -80°C in 20% glycerol and were streaked on 7H11 plates directly from these frozen stocks when required. Single colonies were picked and inoculated routinely from streak plates into 5 ml 7H9 broth with OADC, tween, and the appropriate antibiotic, and incubated standing at 37°C until saturated. *M. tuberculosis* strains constructed are listed in Table 18.

Table 18. *M. tuberculosis* strains.

Strain background	Relevant mutation(s)	Replicating plasmid	Antibiotic resistance	Recombineering substrate used to construct strain
M. tuberculosis H37Rv	Δ*groEL1*:*res-sacB-hyg*^{R}*-res*	pJV53	Kan^R, Hyg^R	AES; pMtbgroEL1
M. tuberculosis mc^{2}7000	*rpoB* H451R	pJV62	Kan^R, RIF^R	ssDNA (JCV326)
M. tuberculosis mc^{2}7000	*rpsL* K43R	pJV62	Kan^R, Str^R	ssDNA (JCV330)

These are strains with specifically engineered mutations that were constructed by recombineering; the type of recombineering substrate used for each mutation is described. This list does not include strains constructed merely by introducing a replicating or integrating plasmid. Abbreviations: AES, allelic exchange substrate; res, resolvase site.

6.9.3.2 Competent cell preparations

Competent cells were prepared as described [229] and similarly to those as described for *M. smegmatis* (section 6.9.2.2) with slight differences. Culture volumes used were no larger than 50 ml per 250 ml bottle, and these were grown standing at 37°C for up to 2 weeks, or until they reached an OD_{600} = 0.8. The cells were not incubated on ice, but prepared at room temperature

by pelleting and washing with 10% glycerol and finally resuspending in 10% glycerol in the same manner as for *M. smegmatis*. The cells were typically not frozen but used immediately for transformation. Extra cell aliquots were frozen and stored at -80°C until use.

6.9.3.3 Transformations

M. tuberculosis cells were transformed using the conditions described for *M. smegmatis* above (section 6.9.2.3 and [229]), except for the following: the cells were never incubated on ice, and the cells were recovered – following transformation – in 7H9 broth supplemented with OADC and tween for 1 – 3 days standing at 37°C. Transformations were plated on selective media and incubated at 37°C for 20 – 30 days.

6.10 RECOMBINEERING PROTOCOLS

6.10.1 Strain growth and media

6.10.1.1 *M. smegmatis*

M. smegmatis mc^2155 recombineering strains were made as described [227-229] by transforming the pJV plasmids into wild type electrocompetent cells and plating on 7H10/Kan media. The transformants were streaked for single isolates, inoculated into 3 ml cultures of 7H9/ADC/tween/Kan, grown shaking at 37°C until saturated, and frozen at -80°C. These were then sub-cultured for growing competent cell batches.

To grow competent cells, recombineering strains were inoculated in 7H9 induction medium (7H9, 0.2% succinate, Kan, and tween) to an OD_{600} = 0.010 – 0.025 approximately 15

hr prior to the desired preparation time and incubated shaking at 37°C. The media was prepared by bringing the 90 ml of 7H9 up to 100 ml with dH_2O, and adding 1 ml of 20% succinate (succinic acid dibasic sodium salt) to a final concentration of 0.2%, and these were grown to an OD_{600} = 0.4 – 0.5, acetamide added to a final concentration of 0.2% (to induce gene expression), and grown for 3 hr shaking at 37°C. The competent cells were then prepared as described above (section 6.9.2.2).

Electrocompetent cells of recombineering strains were transformed with the recombineering substrate as described above (section 6.9.2.3). Cells were recovered in 7H9 with ADC and tween for 4 hr (unless otherwise described) and plated on selective media, always with Kan present in addition to the specific antibiotic required for each recombineering protocol.

6.10.1.2 *M. tuberculosis*

M. tuberculosis H37Rv and *M. tuberculosis* mc^27000 recombineering strains were made as described [227-229]. Plasmids for recombineering were transformed into wild type electrocompetent cells prepared as described above (sections 6.9.2.2 and 6.9.2.3) and plated on 7H11/Kan media (plus pantothenate for mc^27000). The colonies were inoculated into 5 ml cultures of 7H9/ADC/tween/Kan (plus pantothenate for mc^27000), grown standing at 37°C until saturated, and frozen at -80°C. These cultures were then sub-cultured for growing competent cell batches.

Recombineering strains were subcultured into 50 ml of 7H9 induction medium (7H9, 0.2% succinate, Kan, tween, and pantothenate for mc^27000) to an OD_{600} = 0.01 – 0.025 and incubated standing at 37°C for approximately 10 days. Once the cultures reached OD_{600} = 0.45 – 0.50, acetamide was added to a final concentration of 0.2%, and the cells were grown at 37°C overnight (>16 hrs). Electrocompetent cells were prepared as described above (section 6.9.3.2).

The cells were transformed as described above (section 6.9.3.3) with the recombineering substrate. Transformed cells were recovered in 7H9 plus OADC and tween (plus pantothenate for $mc^2$7000) for >16 hr (unless otherwise described) and plated on 7H11/Kan selective media containing antibiotics specific to each protocol below.

6.10.2 Recombineering substrates: synthesis and preparation

6.10.2.1 Gene replacements

To construct substrates for making allelic replacement mutants or "gene knockouts (KOs)," allelic exchange substrates (AESs) or "KO substrates" were constructed for each target gene (see Figure 29). These contain homologous sequences upstream and downstream of the gene and were cloned flanking an antibiotic resistance cassette. Primers were designed to amplify ~500 bp regions of homology at the 5′ and 3′ ends of the gene, typically designed such that ~100 bp at each end of the target gene is intact following gene replacement. Primers also were engineered to contain specific restriction enzyme sites to facilitate directional cloning. The PCR products were cloned into a vector flanking a *hyg*-resistance cassette, typically either pYUB854 (containing γδ res sites for unmarking), pJV69 (pYUB854 without res sites), or pJV150 (pJV69 plus a sacB cassette). The cloning was often performed as a 4-way ligation in which the cloning vector was digested with all four restriction enzymes, corresponding to the sites in the PCR primers, and these two pieces were ligated simultaneously to the two digested PCR products to yield one final plasmid (confirmed by analytical restriction digest).

The vector (containing the homologous sequences) was linearized by restriction digest with two enzymes, preferably the two enzymes used to clone at the most distal regions of the

targeting substrate. This yielded two fragments; one containing the *hyg*^{R} cassette flanked by the two homologous regions, and the other fragment containing the *oriE* backbone. Alternatively, the section with the homologous regions was amplified by PCR (such as for the pMsgroEL1KO substrate). The digest reaction or PCR reaction was cleaned up to remove enzyme using the QIAquick gel extraction protocol for enzymatic cleanup (QIAGEN), and DNA was eluted in dH_2O (in order to minimize salt for transformations). The linear DNA containing the homology was quantified by agarose gel electrophoresis or UV spectrometry.

6.10.2.2 Point mutations

Substrates for making point mutations are ssDNA oligonucleotides. The shortest recommended length is 48 nucleotides (Figure 25; [228]), although longer substrates (70 nt – 100 nt) were used for some experiments. The mutation(s) to be introduced was centered in the oligonucleotide. From experimental evidence it was determined that oligonucleotides that are complementary to the lagging strand for DNA synthesis work better than those that anneal to the leading strand. The oligonucleotides were synthesized as described above (DNA substrates) and resuspended in TE buffer upon receipt of the lyophilized DNA pellet, typically to 1 μM. The sequences of oligonucleotides used in this study are summarized in Table 16.

6.10.2.3 Unmarked deletions

Recommended substrates for deletions are 200 bp, with 100 bp of homology on each side of the deletion locus based on a method previously described [172]. First, a 100 nt oligonucleotide was designed with 50 nt of homology on each end. Primers, called "extenders," were designed that contained 50 nucleotides at the 5′ end that had homology to the target gene

followed by 25 nt that annealed to the template (100 nt oligonucleotide). The final PCR product was 200 bp and contained 100 bp of homology upstream and downstream of the target gene/region (Figure 28).

6.10.3 Construction of mutants

6.10.3.1 Gene replacements

Electrocompetent cells of strains containing plasmid pJV53 (or a similar plasmid containing Che9c *60-61*) were transformed with 100 ng targeting substrate DNA as described above (6.9.2.3 and 6.9.3.3). The transformations were recovered by incubating at 37°C in 7H9 broth containing ADC and tween, and OADC if for *M. tuberculosis*. For *M. smegmatis*, the cells were recovered for 4 hrs, and for *M. tuberculosis*, the cells were recovered 1-3 days.

Following recovery, the entire reaction (~1 ml) was plated on 7H10 or 7H11 agar plates (containing Kan and Hyg, and oleic acid for *M. tuberculosis*), and incubated at 37°C until colonies were of sufficient size for sub-culturing (~5 days for *M. smegmatis*, 3-4 weeks for *M. tuberculosis*). Typically, between 50-200 recombinant colonies were recovered. All batches of competent cells were tested for cell competency by transforming (in a separate aliquot of cells) 50 ng of a Hyg^R integration-proficient vector (either pJV39 or pSJ25Hyg), plating on 7H10/Hyg plates, and determining the number of cfu per μg of plasmid DNA. The number of viable cells in each transformation reaction was determined by plating serial dilutions of the cell competency control reaction on 7H10/Kan media.

6.10.3.2 Point mutations

Selectable point mutations in either the mycobacterial chromosome or on extrachromosomal plasmids were generated by transforming the ssDNA substrate (containing the point mutation) into electrocompetent recombineering cells. The strain background was typically pJV62 or a derivative, since only the recombinase (gp61) is required for recombination with ssDNA. 100 ng of ssDNA substrate was transformed as described above (6.9.2.3 and 6.9.3.3), and the cells were recovered in 7H9 broth with ADC, tween, and OADC for *M. tuberculosis*, at 37°C. For *M. smegmatis*, cultures were recovered shaking for 4 hrs, and for *M. tuberculosis* they were recovered for 3 days standing. The recovered cells were diluted, plated on selective media, and incubated for 4-5 days (*M. smegmatis*) or 3 weeks (*M. tuberculosis*) at 37°C. Cell competency and viability counts were determined as described above for gene replacements.

For non-selectable mutations, transformations with ssDNAs were performed as above using excess ssDNA compared with a Hyg^R, integration-proficient plasmid (pJV39 or pSJ25Hyg) or a ssDNA that could also be recombined to confer Hyg^R (JCV198 ssDNA in backgrounds with hyg^S). Optimal results were obtained with 500 ng of the mutating ssDNA and 100 ng of the Hyg^R selectable element, respectively. Following recovery (4 hr for *M. smegmatis*, 3 days for *M. tuberculosis*), the cells were diluted in 7H9 broth (plus ADC/OADC and tween) containing Hyg and Kan to approximately 10–100 Hyg^R cells per well (1 ml media) in a sterile 96-well culture block. These dilutions were simultaneously plated on 7H10/Hyg/Kan agar plates to determine Hyg^R cell counts. The cultures were incubated (shaking at 250 rpm for *M. smegmatis*; standing for *M. tuberculosis*) at 37°C to an OD_{600} = 1.0 and screened by colony PCR (MAMA-PCR). Each culture well containing a mutant allele was plated for single colonies and re-screened by colony PCR to identify the isolated mutant.

6.10.3.3 Unmarked deletions

To make unmarked deletions, 100 – 200 ng of the dsDNA substrate was co-transformed with 50 – 100 ng of a Hyg^R-selectable substrate, either a Hyg^R plasmid or another ssDNA that could confer Hyg^R (JCV198 in a *hyg*^{S} background strain). The cells were recovered as described for non-selectable point mutations; the transformation was either plated on Hyg media or were serially diluted in 7H9 broth with Hyg. Transformants or liquid culture dilutions were screened by PCR or identifying phenotype (*e.g.* leucine auxotrophy for *leuD* mutants) for the presence of the desired mutation. Each culture containing the mutation was plated for single colonies and re-screened by PCR to identify the mutant strain.

6.10.4 Analysis of recombinant colonies

6.10.4.1 Gene replacements

Colonies recovered from transformations with the targeting substrate were analyzed by either colony PCR or Southern blot to confirm the genotype of the strain. For colony PCR, the colonies were either allowed to grow to a large size or were patched onto a fresh plate to get enough cells for the PCR. Primers were designed within the homologous regions of the targeting substrate to determine if the gene locus contained the Hyg^R resistance cassette or if it was wild type, which would result in differently sized PCR products. Colony PCR was performed as described above (6.3.1).

For Southern blot analysis, colonies were inoculated into ~10 ml 7H9 broth containing ADC, tween, Kan, and Hyg (and OADC for *M. tuberculosis*). The cultures were incubated at 37°C for 3 days (*M. smegmatis*) or 10 days (*M. tuberculosis*) until the culture had a substantial amount of visible growth. The cells were collected and genomic DNA was prepared as described

above (6.8.1). Southern blot analysis was performed as described above (6.8) on each recombinant strain to be tested, using the pJV53 strain as a control. Probes were synthesized using primers either to the upstream or downstream homologous region in order to determine if the gene locus contained the Hyg^R resistance cassette or if it was wild type. Alternatively, a probe to the Hyg^R cassette was also used.

6.10.4.2 Point mutations

Point mutations that were selectable (such as the *rpsL* K43R mutation) were confirmed by sequencing. PCR was performed with primers that amplified the gene locus, and this DNA was cleaned up using QIAquick PCR purification (QIAGEN) and sequenced as described above using the same primers.

For large-scale identification of non-selectable point mutations, culture wells or single colonies were screened by MAMA-PCR with primers designed to distinguish between wild type and mutant alleles as described above (6.3.2) [30,219]. Primers were synthesized for both the mutant allele and the wild type allele as a control. Positive control templates for mutant alleles were synthesized by PCR-amplification by one of four methods: 1) using the recombineering substrate that was used to make the point mutation as a forward primer, 2) synthesizing a new primer that contained the mutation as a forward primer, 3) using the mutant MAMA-PCR screening primer to amplify from a wild type template but with the *Pfu* polymerase that could read through the non-matching 3′ end of the wild type template, or 4) PCR amplifying a known mutant with sequencing primers.

6.10.4.3 Unmarked deletions

Deletion mutants were identified by PCR from a colony or culture using primers that flanked the deleted region, which were designed to amplify either wild type or mutant loci. The deletion mutants would therefore yield a smaller-sized PCR product. Colony PCR was performed as described above (6.3.1), and cultures that contained mutant alleles were plated for single colonies and re-screened to identify the isolated mutant strain.

6.10.5 Strain unmarking

Gene replacement strains that were constructed by recombineering often contained γδ-res sites for removing the interrupting Hyg^R cassette (if the targeting substrate was constructed in a pYUB854 vector backbone). Alternatively, dsDNA recombineering was used to remove the Hyg^R cassette, but only in conjunction with a *sacB* cassette for negative selection (if the targeting substrate was constructed in a pJV150 vector backbone). Ultimately, the pJV53 plasmid (or similar recombineering plasmid) was occasionally removed from the strain by serial dilution, or by using *sacB* as a negative selection in strains that contain a pJV recombineering plasmid with a *sacB* cassette (*e.g.* pJV48, pJV126).

6.10.5.1 Removing Hyg^R by γδ-resolvase

Unmarking of *M. smegmatis* recombinant replacement strains using γδ resolvase was accomplished by transforming the strain with pGH542, a Tet^R plasmid that constitutively expresses the resolvase. Electrocompetent cells of the recombinant strain were prepared, transformed with 50 ng of plasmid pGH542, and plated on 7H10 agar plates containing Tet. The plates are incubated at 37°C for 3-4 days until colonies are of sufficient size for sub-culturing.

The recovered colonies were patched onto multiple selective 7H10 agar plates containing antibiotics in the order as follows: 1) Hyg, 2) Tet, 3) Cb/Chx, and were incubated at 37°C for 3-4 days. Colonies that are Hyg^S and Tet^R are therefore "unmarked" and the removal of the Hyg^R cassette can be verified by colony PCR at that locus.

The pGH542 plasmid was removed from this strain at the same time as the recombineering plasmid following the protocol described below (6.10.5.2). However, occasionally it was desired to retain the recombineering plasmid but remove the pGH542 plasmid, and in these cases, the recombineering plasmid was selected during the serial dilutions, whereas the pGH542 was not (media containing Kan and not Tet). Ultimately, strains that were determined (by patching on multiple selective plates) to be Tet^S, Kan^S, and Hyg^S were retained and stored.

6.10.5.2 Removing the recombineering plasmid

In cases in which it was desired to remove the recombineering plasmid from a recombinant *M. smegmatis* strain, this was subcultured into a 10 ml culture of 7H9 media (containing only ADC, tween, and Cb and Chx) and incubated with shaking at 37°C until it reached saturation (~2 days). This culture was subcultured into another 10 ml culture exactly as above, using 1 μl of the culture (1:10,000) and incubated with shaking at 37°C until saturation (~2 days). Dilutions (10^{-4} – 10^{-7}) of this culture were plated on 7H10 agar plates (containing Cb and Chx only). The recovered colonies were patched onto multiple selective 7H10 agar plates containing antibiotics in the order as follows: 1) Hyg, 2) Tet, 3) Kan, 4) Cb/Chx only. The plates were incubated at 37°C for 3-4 days. Kan^S colonies were saved and stored.

6.11 MYCOBACTERIOPHAGE MANIPULATIONS

6.11.1 Mycobacteriophage lysate preparation

Mycobacteriophages TM4 and Che9c were propagated on *M. smegmatis* mc^2155 as described [198]. *M. smegmatis* cultures were grown in ADC (no tween) until saturated in baffled flasks. For typical small-plate lawns, 1.5 ml MBTA was mixed with 1.5 ml 7H9/ADC/$CaCl_2$ and 300 μl *M. smegmatis* cultures, and this was solidified and used for spot-tests with serially-diluted phage. For plate infections, serially-diluted phage were added to the 300 μl of cells, incubated standing at 37°C for 20 mins and subsequently plated with MBTA and 7H9 as above. Lysates were prepared by flooding plates (5-8 ml) with phage buffer plus $CaCl_2$ at room temperature for 2 hrs, or overnight at 4°C. These were collected, the debris removed by centrifugation at 3500 x *g*, and the supernatant was filtered (0.22 μM filters) and stored at 4°C. For large-plates, 5 ml MBTA, 5 ml 7H9/ADC/$CaCl_2$ and 1 ml *M. smegmatis* cells were used.

6.11.1.1 Large-scale preparation of mycobacteriophage CsCl stock

TM4 was prepared in which 30 large plates were made of TM4 infections that yielded "webbed" lawns (approximately 6000 pfu total per large plate). Lysates were prepared, and these were treated either by the conventional PEG precipitation protocol and continuous CsCl gradient purification [198], or using the following protocol.

For the modified TM4 “large prep protocol,” the lysates collected from 30 large plates were centrifuged in four Ti75 rotor tubes (~60 ml per tube) at 20,000 x *g* in the ultracentrifuge for 1.5 hr (phage buffer was used to balance tubes). The supernatant was removed, and the phage pellet was resuspended in 2 ml phage buffer (plus 1 mM $CaCl_2$) overnight standing at 4°C. The phages were collected by gently swirling the pellet, and this was titered. At this stage, either a continuous CsCl or step CsCl gradient was used to purify the phage from any cell debris. The step gradient yielded better results, and this was performed by layering the following solutions in this order: 1) 8 ml phage into one tube, 8 ml phage buffer in another tube (as a balance), 2) Pasteur pipeting 1 ml 10% glycerol under phage/phage buffer, 3) 1.5 ml of 1.4 mg/ml CsCl under glycerol, and 4) 1.5 ml of 1.6 mg/ml CsCl under that. These tubes were centrifuged at 30,000 x *g* in a swinging bucket SW41 rotor for 1.5 hrs. The phage band was extracted with a syringe, and the phage were dialyzed twice against 500 ml phage buffer with 1 mM $CaCl_2$ at 4°C for 4 hr each time.

6.11.1.2 Genomic DNA isolation from mycobacteriophage stock

Genomic DNA was isolated as described [198]. Briefly, ~500 µl of dialyzed phage were extracted with buffer-equilibrated phenol repeatedly (with TE back-extraction) until protein was removed. This was followed by an extraction with phenol:chloroform:isoamyl alcohol (25:24:1) and another extraction with chloroform. The DNA was ethanol precipitated, washed, and resuspended in TE.

6.11.1.3 Small-scale genomic DNA isolation from lysates

To make small amounts of phage DNA, 2-3 small plate lysates were prepared, and the phage was pelleted by incubating on ice with equal volumes of saturated ammonium sulfate for

1-3 hr [198]. The phage were pelleted at 3500 x *g* and resuspended in phage lysis buffer. This was treated with proteinase K at 10 μg/ml for 2 hr at 37°C and subsequently phenol extracted as described above. It is important to note that the aqueous layer in the first extraction is underneath the organic layer due to the presence of the lysis buffer. To fix this, an equal volume of TE was added to the phage lysis buffer following treatment with proteinase K; this resulted in a shift of the aqueous layer to the top for all extractions.

6.11.2 TM4 Cosmid library construction

Approximately 4 μg TM4 genomic DNA was ligated to itself to form concatemers using T4 DNA ligase overnight at room temperature; this facilitated connection of the otherwise linear ends. This was partially digested with a frequently-cutting enzyme to yield approximately 40-45 kbp fragments; this was accomplished best with a 15 min digest with *Sau3A*I at 37°C. This was immediately placed on ice, phenol extracted, ethanol precipitated, and resuspended in 10 μl TE. The DNA was then ligated to digested pYUB854 DNA cut with *Bgl* II (ends compatible with *Sau3A*I) using the FastLink ligase (Epicentre) for 15 mins at room temperature, and packaged into λ heads using Gigapack III Gold Packaging Extract (Stratagene) according to manufacturer's instructions. HB101 cells were grown and infected with the λ-packaged molecules, and colonies were selected on Hyg (resistance conferred by pYUB854). Colonies were miniprepped, and these were analyzed by restriction digest and sequencing to identify the segment of TM4 that was cloned into the cosmid. Alternatively, pools of *E. coli* HB101 colonies were prepared together by scraping colonies into 25 ml LB broth plus Hyg, growing to saturation, and midi-prepping the DNA (QIAGEN).

6.11.3 TM4 cosmid recombination assays

TM4 cosmids that were defined by analytical restriction digest and sequencing were used to assay for recombination *in vivo* in *M. smegmatis*. Individual cosmids or pooled cosmid preps were transformed into electrocompetent *M. smegmatis* as described above (6.9.2.2 and 6.9.2.3), except tween was not used in the cultures. Concentrated stocks of cosmid DNAs were critical for good transformation frequencies; 1 μg total DNA was transformed (500 ng each cosmid if in pairs), and 50-200 ng TM4 DNA was used as a control. Transformations were recovered in 1 ml LB broth for 30 min and plated in 300 μl *M. smegmatis* cells with 0.5 ml 7H9 and 1.5 ml MBTA on 7H10 agar plates. These were incubated at 37°C overnight, and plaque numbers recorded for single cosmid transformations versus pair-wise combinations. PCR assays to detect TM4 phage DNA (band = 880 bp) and pYUB854 DNA (band = 584 bp) simultaneously contained primers to both (TM41444-1464 880F, TM42323-2303 880R; pYUB854 509-533F, pYUB854 1069-1093R) and were conducted as described above (6.3).

APPENDIX A

THE ROLE OF HOST NUCLEASES IN MYCOBACTERIAL RECOMBINEERING

A.1 INTRODUCTION

A.1.1 The *E. coli* RecBCD complex and λ Gam

The RecBCD complex in *E. coli* functions for recombinational repair of double-strand DNA breaks and broken replication forks (reviewed by A. Kuzminov [106] and by Amundsen and Smith [5]). It is a highly processive multienzyme complex that possesses strong helicase activity and ATP-dependent dsDNA and ssDNA exonuclease activities. During repair, the enzyme degrades dsDNA in a 5′ to 3′ direction, and following recognition of a Chi site (by RecD), RecBCD stimulates RecA polymerization on the 3′ ssDNA tail (along with SSB), which promotes strand invasion at homologous targets for recombination. However, the RecBCD complex also degrades foreign dsDNA molecules – such as the ends of the λ genome – after digestion by restriction enzymes upon entry into the cell. More than 40 years of research have been dedicated to examining RecBCD, providing a detailed understanding of the genetic and

biochemical properties of this very complex enzyme [106]. However, this appendix will focus on a small portion of this work and discuss only the most pertinent aspects of its biology.

Mutant *E. coli* strains of *recB, recC,* and *recD* display different phenotypes in response to DNA damage induced by UV, as well as in regard to their recombination activity [106]. Null *recB* or *recC* mutations render cells sensitive to UV and have decreased overall viability (~30%), whereas *recD* mutants survive like wild type after UV treatment [133]. The UV phenotype of *recBCD* mutant strains can be suppressed by expression of λ Exo/Beta, and the viability of this strain is even increased ~10-fold compared to wild type [135]. Further, the nuclease activities of RecBCD are eliminated in a either a *recB* or *recD* mutant strain [5]. However, in a *recD* mutant strain, some of the helicase activity is retained, although it is decreased because RecD provides the faster helicase subunit [5]. These *recD* strains are described as 'hyper-recombination' mutants [106] because they constitutively load RecA for homologous recombination. This phenotype essentially mimics the activity of RecBCD following recognition of Chi, which is attributed to the removal of the RecD subunit after Chi; the presence of RecD can actually inhibit RecA polymerization [5]. Interestingly, there is a specific *recB* mutation (*recB1080*) that abolishes nuclease activity but can still unwind DNA; however, this is incapable of recombination due to the presence of RecD [5].

The λ Gam protein inhibits all known activities of RecBCD by binding to the RecB subunit and preventing the complex from binding to dsDNA ends [39,93,122,133,138,176]. However, Gam expression does not cause all the examined phenotypes of a Δ*recBCD* strain. This is due to the presence of a portion of RecBCD that is not bound by Gam and is therefore able to degrade dsDNA [138]. There is also some evidence that Gam interacts with SbcC, though this is not well studied [102]. Because of these activities, dsDNA is protected from nuclease

attack in the presence of Gam, which makes λ Red recombination more efficient [38,142,240,241]. Therefore, although Gam is not essential for phage λ propagation [53] or recombineering [43,241], expression of Gam in recombineering assays with dsDNA increases frequencies ~10-fold [43]. Gam expression in wild type cells yields a UV-sensitive phenotype that mimics a *recBC* strain (~10-fold decrease in viability compared to wild type) [135]. However, Gam does not increase the sensitivity of either *recBC* or *recD* strains, which suggests that the UV phenotype is a result of Gam interaction only with RecBCD and not any other complex [135].

Other phages encode proteins that function analogously to λ Gam, such that they protect linear dsDNA ends from degradation, although the means by which they accomplish this differs. For example, both phage Mu (Gam) [2] and phage T4 (gp2) [6,115,208] have proteins that bind dsDNA ends, whereas λ Gam specifically inactivates RecBCD. Conversely, the Abc proteins of P22 actually modify the activity of RecBCD in order to exploit its 5′-3′ ssDNA exonuclease activity for P22 Erf-mediated recombination [136,176]. A decrease in host nuclease activity has been observed following infection with several additional dsDNA phages [195], which is most likely being facilitated by similar types of proteins.

A.1.2 Mycobacteria encode both RecBCD and AddA homologues.

Recent bioinformatic analysis of several mycobacterial genomes has revealed a number of genes that encode homologues of the *B. subtilis* AddA protein (Figure 34, D. Ennis and G. Cromie, personal communication). *B. subtilis* AddA is a part of the AddAB enzyme complex. These proteins are functionally analogous to *E. coli* RecBCD *in vivo* [98], although the complexes have slightly different biochemical properties. These genes were originally believed to be restricted to

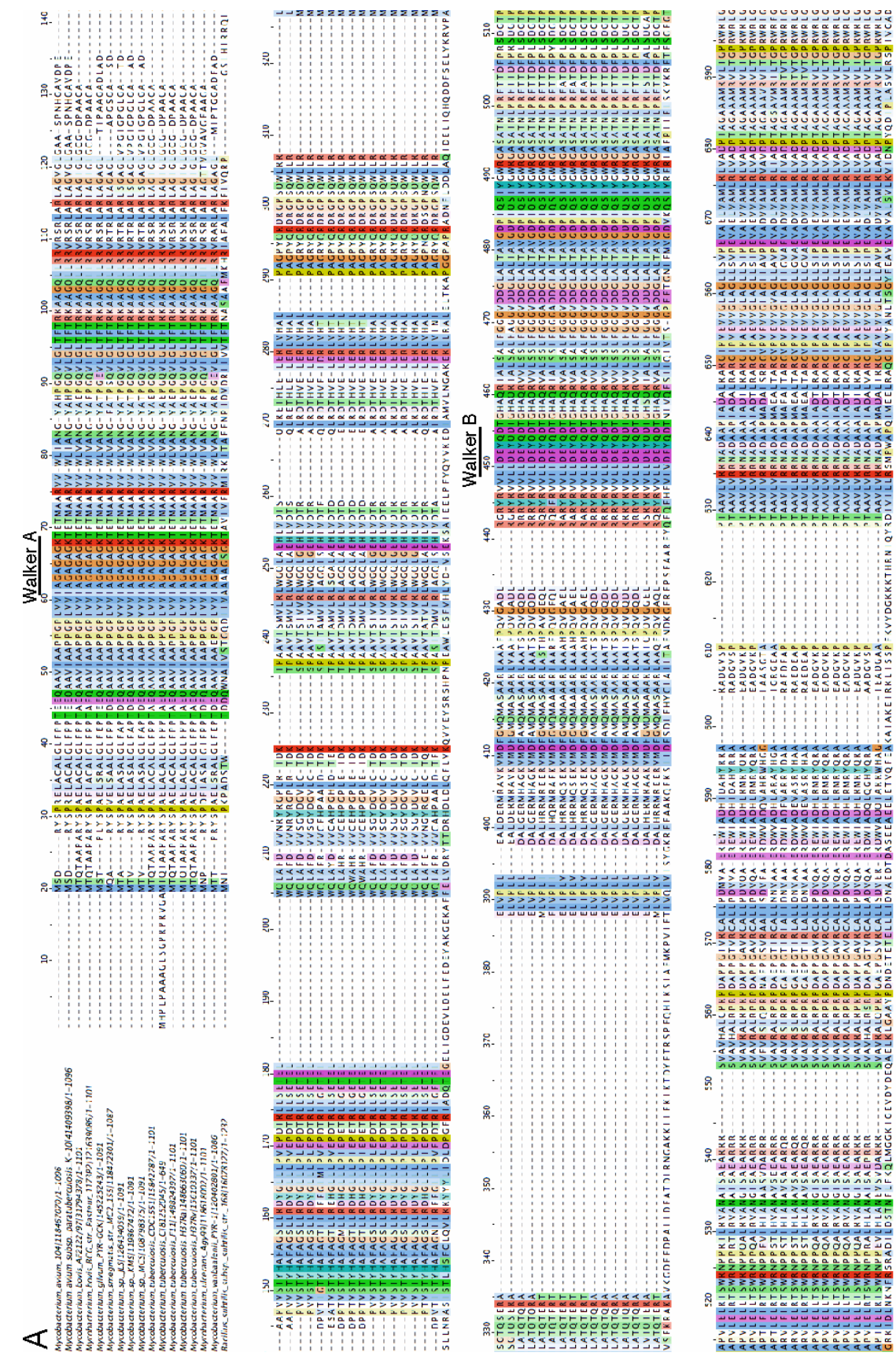
Gram-positive bacteria [32], but recently they have also been identified in the Gram-negative bacteria, *Rhizobium etli* [245] and *Coxiella burnetii* (D. Ennis, personal communication). The AddA subunit has homology to RecB and contains a similar nuclease domain, as well as helicase and ATPase domains. The AddB subunit does not have homology to RecC or RecD, but it does have ATPase and nuclease domains that are slightly divergent in sequence compared to AddA [32].

It is apparent that the list of genomes that encode these proteins is not likely complete, and also that some bacteria may encode both types of enzymes. For example, *M. tuberculosis* was listed in the review by Chedin *et al.* as having only a 'RecBCD-type enzyme' [32], but bioinformatic data suggests that AddA homologues are also present. Further, there is an *E. coli* RecD homologue present in both *B. subtilis* and *Lactococcus lactis*, as well as in other bacteria that contain neither RecBC nor AddAB. Collectively, these findings suggest that there is still much to learn about the roles of these proteins in bacteria that encode homologues of both types of enzymes.

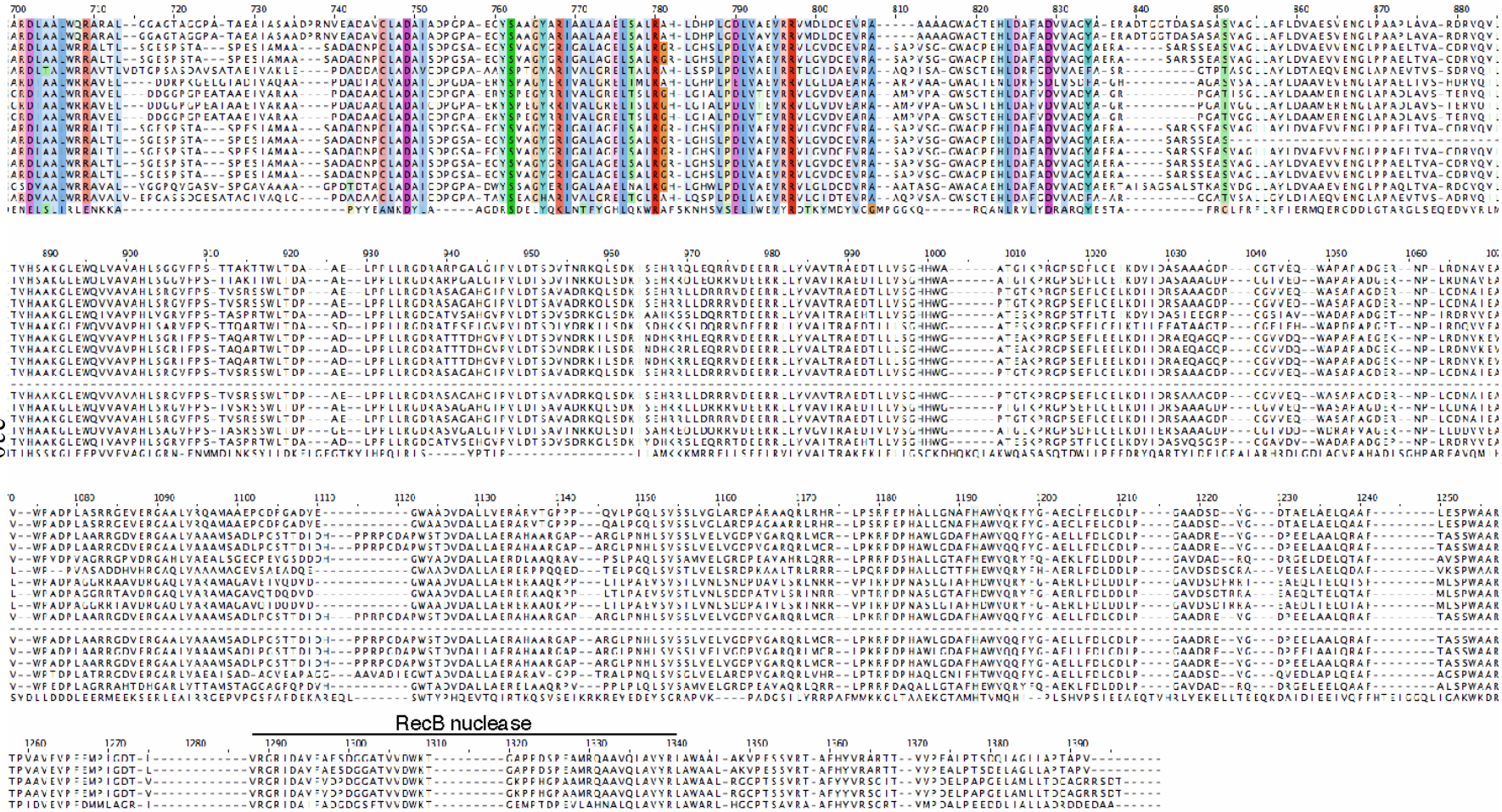
The putative mycobacterial AddA proteins are highly conserved between species (>60% amino acid identity) and are similar to the *B. subtilis* AddA protein at specific regions, including the ATPase domains (Walker A and B motifs) and RecB nuclease domains (Figure 34). For example, *M. smegmatis* MSMEG_1943 has 24% and 27% identity to *B. subtilis* AddA at its N- and C-terminus, respectively. However, BLAST analysis with *B. subtilis* AddB does not identify any obvious homologues in the mycobacteria. Instead, these analyses identify multiple genes in each mycobacterial genome with similarity to AddA, and some of these are found in pairs (*e.g.* MSMEG_1943 and MSMEG_1941). Typically, the gene with the highest degree of identity was aligned in Figure 34A, and the adjacent gene was aligned in Figure 34B.

The high degree of sequence similarity between the mycobacterial AddA-like ORFs and *B. subtilis* AddA suggests that these bacteria encode not only RecBCD, but also an AddA nuclease. However, it is unclear if the mycobacterial AddA proteins are active or if they function in a complex with another protein, such as the more distantly related AddA-like proteins encoded by the adjacent genes. It has yet to be determined, to my knowledge, if these genes are expressed *in vivo* or active in any mycobacterial species. It will be interesting to observe if the AddA and RecBCD proteins are functionally redundant, or if one affords a separate function *in vivo.*

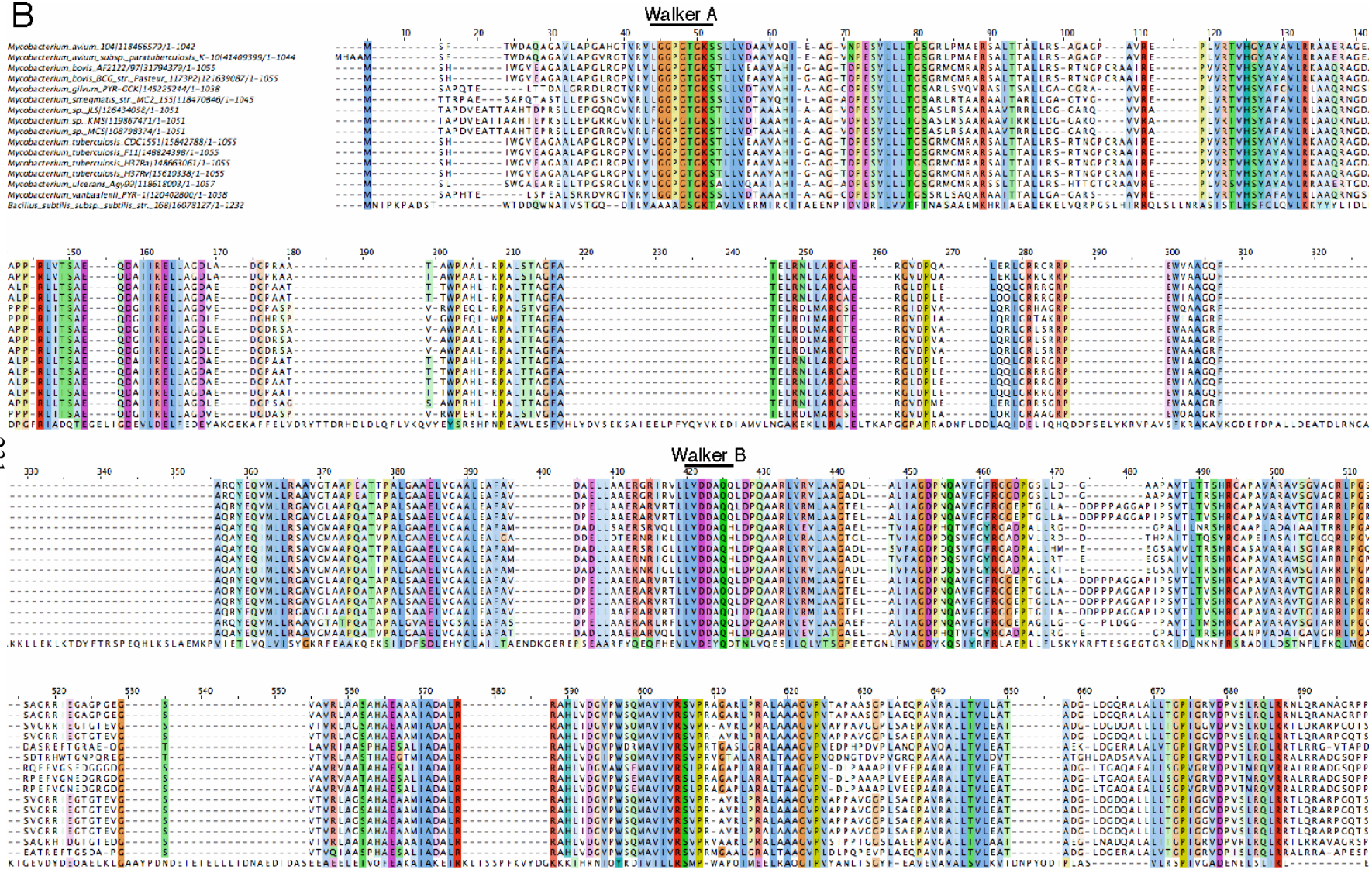
Figure 34. Multiple sequence alignments of putative mycobacterial and *B. subtilis* AddA proteins.



A (continued)



RecB nuclease


B
Walker A
Walker B
Mycobacterium_avium_104|118466573/1-1042
Mycobacterium_avium_subsp._paratuberculosis_K-10|41409399/1-1044
Mycobacterium_bovis_AF2122/97|31794373/1-1055
Mycobacterium_bovis_BCG_str._Pasteur_1173P2|121639087/1-1055
Mycobacterium_gilvum_PYR-GCK|145225244/1-1038
Mycobacterium_smegmatis_str._MC2_155|118410846/1-1045
Mycobacterium_sp._JLS|126434058/1-1051
Mycobacterium_sp._KMS|119867471/1-1051
Mycobacterium_sp._MCS|108798374/1-1051
Mycobacterium_tuberculosis_CDC1551|15842788/1-1055
Mycobacterium_tuberculosis_F11|148824398/1-1055
Mycobacterium_tuberculosis_H37Ra|148661061/1-1055
Mycobacterium_tuberculosis_H37Rv|15610338/1-1055
Mycobacterium_ulcerans_Agy99|118618003/1-1057
Mycobacterium_vanbaalenii_PYR-1|120402800/1-1038
Bacillus_subtilis_subsp._subtilis_str._168|16078127/1-1232

B (continued)

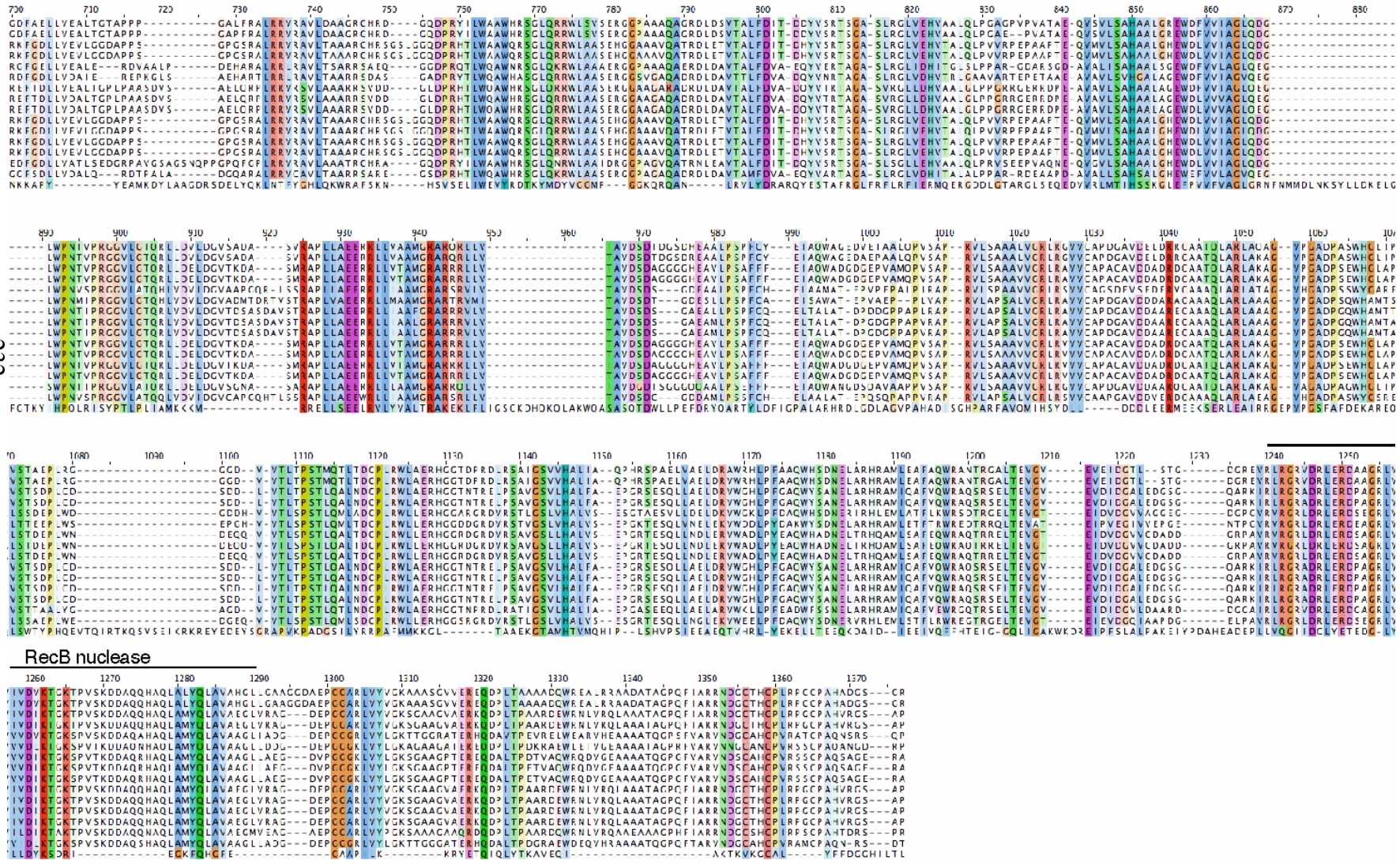


Figure 34. These alignments were performed and provided courtesy of Gareth Cromie. The *B. subtilis* subsp. *subtilis* (bottom line) AddA protein was aligned with predicted proteins from various mycobacterial genomes. Mycobacterial ORFs with similarity to *B. subtilis* AddA were identified and aligned; those with the most similarity were placed in alignment (A) and the adjacent gene with less similarity was aligned in (B). The putative Walker A and B motifs are indicated, as well as the regions with similarity to the RecB nuclease domain. Aligned from top to bottom: *M. avium* 104, *M. avium* subsp. *paratuberculosis* K-10, *M. bovis* AF2122, *M. bovis* BCG str. Pasteur, *Mycobacterium gilvum*, *M. smegmatis* mc^2155, *Mycobacterium sp. JLS*, *Mycobacterium sp. KMS*, *Mycobacterium sp. MCS*, *M. tuberculosis* CDC1551, *M. tuberculosis* C, *M. tuberculosis* F11, *M. tuberculosis* H37Ra, *M. tuberculosis* H37Rv, *Mycobacterium ulcerans* Agy99, *Mycobacterium vabbaalenii* PYR-1, and *B. subtilis* subsp. *subtilis* str. 168.

A.1.3 Recombineering in *E. coli*: the effect of host RecBCD

The degree to which λ Gam is required to inactivate host nuclease activity for λ Red recombineering with dsDNA substrates differs in varying reports. In studies by K. Murphy and Yu *et al.*, the data indicated that Gam expression (or inactivation of *recBCD* by mutation) was absolutely required for recovery of gene replacement mutants [135,240]. This was observed using long (~1 kbp) dsDNA substrates, with either 50 bp or 1 kbp homology lengths. However, later studies using similar length or shorter substrates showed only an ~10-fold decrease in efficiency in the absence of Gam [43,241]. There does not appear to be a difference in recombineering frequencies between a Δ*recBCD* strain or Gam-expressing strain [135], but there are clearly advantages to using Gam to facilitate recombineering in any background. Finally, there is only a modest decrease (~5-fold) in ssDNA recombineering frequencies in the absence of Gam [52].

This appendix will discuss the results of the preliminary experiments that have been performed in *M. smegmatis ΔrecB* and *ΔrecD* strains, as well as assays in which the expression of the λ *gam* gene is controlled by the acetamidase promoter in *M. smegmatis*. These strains were assayed for both UV sensitivity and dsDNA recombineering activity. While these experiments have provided some data regarding the activity of *M. smegmatis* RecBCD, additional experiments will be required to fully understand the role of the mycobacterial nucleases.

A.2 RECOMBINEERING ACTIVITY IN *REC⁻ M. SMEGMATIS* STRAINS

A.2.1 Recombineering in Δ*recB* and Δ*recD* strains

The mycobacteria encode both RecBCD and AddA proteins, as discussed above. The RecBCD genes are grouped together, likely in one operon, in the chromosome of *M. smegmatis* (Figure 35A), the organism in which all these experiments were performed. Conceivably, one approach to increase recombineering frequencies is to inhibit the potentially negative effects of host nucleases. With regard to RecBCD, a *recD* mutant would be ideal because, at least in *E. coli*, it has wild type viability, does not lose viability after treatment with DNA damaging agents, and retains the helicase and recombination-stimulating activities of the complex [5,106]. In order to study the role of RecBCD more thoroughly, *M. smegmatis* mutant strains were constructed that contained gene replacements of either the *recB* and *recD* genes, and were subsequently unmarked by γδ resolvase (Figure 35B,C). These strains were then tested for dsDNA recombineering by targeting the *M. smegmatis groEL1* gene (section 3.3.3), and the data were compared to recombineering frequencies obtained in a wild type genetic background.

The Δ*recB* strain demonstrated a slight increase in recombineering activity, typically between 3- and 5-fold compared to wild type (Figure 35D, Table 19). Since reports in *E. coli* vary regarding the effect of RecBCD, it is difficult to directly compare these data. However, it was surprising that the increase was only 5-fold, particularly because the smallest effect observed in *E. coli* is a 10-fold increase in efficiency [43]. The Δ*recD* strain was also slightly increased for recombineering activity, although to a lesser extent than Δ*recB* (~2-fold increase). These results

indicate that while *recB* and *recD* may have a small effect on recombineering efficiency, their activities do not drastically inhibit recombination with dsDNA substrates.

Figure 35. Recombineering frequencies in Δ*recB* and Δ*recD M. smegmatis* strains.

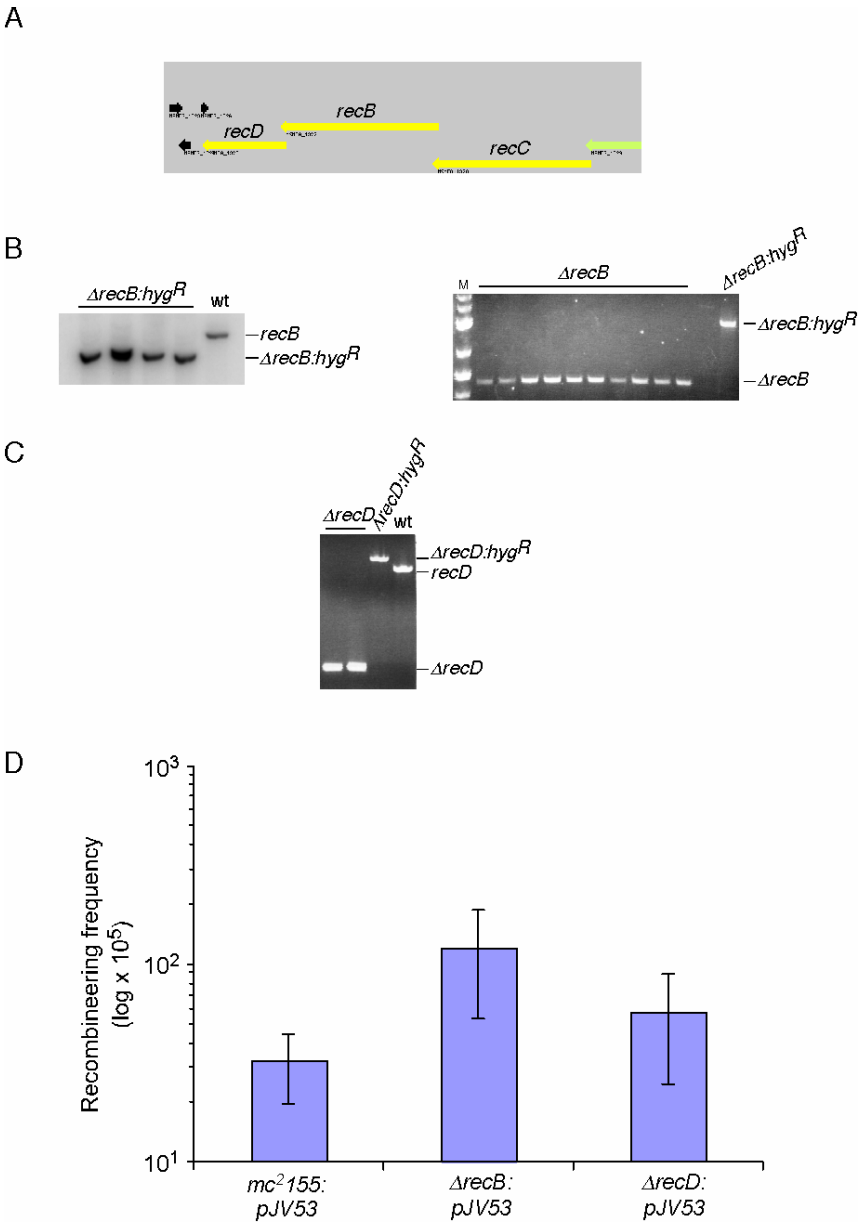


Figure 35. (A) Schematic of the *M. smegmatis* chromosome at the region encoding the *recBCD* genes. (B) Southern blot analysis of DNA prepared from four individual *recB* gene replacement mutants made by recombineering; wild type DNA is used as a control (4,077 bp and 5,175 bp, respectively). Also shown, colony PCR analysis of 10 Δ*recB* mutants that were unmarked with γδ resolvase; the *hyg*^{R} marked mutant strain is used as a control (1,396 bp and 3,130 bp, respectively). (C) Colony PCR analysis of both unmarked and marked *recD* gene replacement mutants made by recombineering, using wild type DNA as a control (401 bp, 2,138 bp, and 1,787 bp, respectively). (D) Recombineering frequencies from experiments targeting the *M. smegmatis groEL1* gene in wild type, Δ*recB*, and Δ*recD M. smegmatis* mc^{2}155 strains containing plasmid pJV53. Frequencies are represented on a log scale and multiplied by 10^5 for presentation purposes. The data shown are the averages of three experiments for wild type and Δ*recB*, and two experiments for Δ*recD*; error bars represent standard deviation.

A.2.2 Expression of λ gam in *M. smegmatis*

In order to determine the activity of λ Gam in mycobacteria, this gene was cloned either singly under $P_{acetamidase}$ (pJV99) or P_{hsp60} (pJV97) or together with the Che9c genes *60* and *61* under $P_{acetamidase}$ (pJV98). In the latter case, the *gam* gene plus 19 bp of sequence upstream of the start codon were cloned downstream (158 bp) of Che9c *61* in plasmid pJV53. Competent cells of an *M. smegmatis* strain containing the pJV98 plasmid (Che9c *60* and *61*, λ *gam*) were prepared similarly to Che9c recombineering strains and tested for dsDNA recombineering activity using *groEL1* as a target gene (sections 6.10.1.1 and 3.3.3). The results did not show a difference in recombineering frequency as compared to the typical pJV53 strain that expresses only Che9c genes (Table 19). This may be due to expression problems, protein instability, or potentially the Gam protein is inactive in *M. smegmatis.*

Table 19. Recombineering frequencies in Δ*recB*, Δ*recD*, and Gam-expressing *M. smegmatis* strains

Strain (proteins encoded)[a]	Recombinant colonies[b]	Cell competency[c]	Recombineering frequency[c]	Ratio[e]
mc^2155:pJV53 (Che9c gp60/61)	226	6.0×10^6	3.8×10^{-4}	N/A
mc^2155:pJV53 Δ*recB* (Che9c gp60/61)	254	1.7×10^6	1.5×10^{-3}	3.9
mc^2155: pJV53 Δ*recD* (Che9c gp60/61)	30	5.3×10^5	5.6×10^{-4}	1.5
mc^2155:pJV98 (Che9c gp60/61, λ Gam)	123	3.7×10^6	3.3×10^{-4}	0.9

a. The *M. smegmatis ΔrecA* strain was constructed by allelic gene replacement with recombineering and unmarked using γδ resolvase, as described in the Materials and Methods.
b. Electrocompetent cells of the strains were transformed with 100 ng of the *groEL1* AES (see Figure 18), and Hyg^R colonies were recovered; the data represent the average of two experiments.
c. Cell competency is determined as the cfu/μg plasmid pJV39, an integration-proficient vector providing hygromycin resistance, when 50 ng was transformed.
d. Recombineering frequency is calculated as the number of recombinant cfu per μg DNA divided by the cell competency.
e. The ratio is calculated by dividing the recombineering frequency of the test strain (Δ*recB*, Δ*recD*, or pJV98) by the recombineering frequency of the wild type mc^2155:pJV53 strain.

A.3 EXAMINATION OF THE UV PHENOTYPE OF *M. SMEGMATIS* STRAINS

The experiments performed with the various *M. smegmatis* strains (Δ*recB*, Δ*recD*, or Gam-expressing) suggested, somewhat inconclusively, that RecBCD has a minor effect (if at all) on recombineering frequencies. To further assess the activity of RecBCD, and specifically to determine the effect of expressing Gam, these strains were subjected to UV treatment to assay for sensitivity to DNA damage. As expected, the Δ*recA* strain consistently showed a decrease in viability that ranged from 100- to 1,000-fold (depending on the assay; Figure 36 and data not shown). However, the Δ*recA* percent survival is higher (100-fold) in this experiment and others compared to previous studies with this strain (~0.1% after treatment with 39 J/m^2; [155]).

Surprisingly, even with a high level of UV treatment (300 J/m^2), the Δ*recB* strain showed viability similar to wild type (Figure 36). Two different Δ*recB* strains have been constructed (*mc*^{2}155:pJV53 Δ*recB*, Figure 36, and mc^{2}155 Δ*recB*, a gift from K. Derbyshire), and both demonstrate wild type viability following UV treatment (data not shown). Further, the Δ*recD* strain also showed wild type viability, but this is similar to what is observed in *E. coli* for a *recD* null mutant [133].

With regard to λ Gam, the mc^{2}155:pJV99 strain ($P_{acetamidase}$*:gam*) appeared to have a slight viability decrease (3.5-fold) compared to the control (mc^{2}155:pJV96) in this particular experiment. However, repetition of this assay did not show the same defect. Further, a similar strain in which *gam* is expressed from the constitutive *hsp60* promoter did not show a decrease in viability. These assays suggest that although Δ*recB* strains are increased in recombination, they are not UV-sensitive. In addition, the strains expressing λ Gam do not appear to have either a UV or recombination phenotype.

Figure 36. UV phenotypes of Δ*recA*, Δ*recB*, Δ*recD*, and Gam-expressing *M. smegmatis* strains

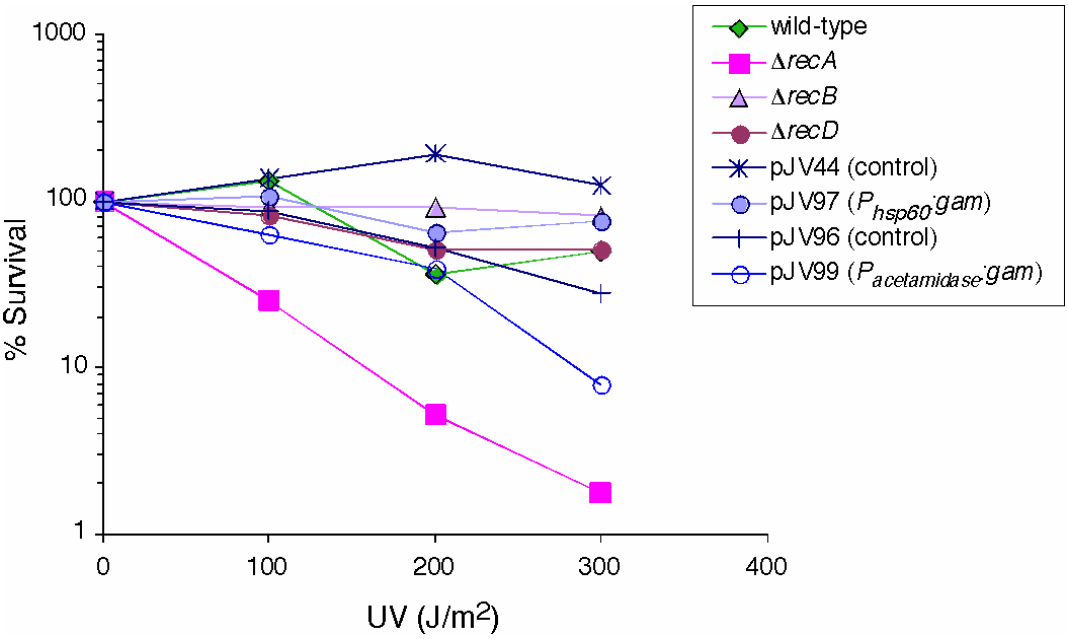


Figure 36. *M. smegmatis* strains were assayed for UV sensitivity by exposure to 100, 200, and 300 J/m^2 UV light and plated on solid media to determine the number of viable cells as described in the Materials and Methods (section 6.9.2.4). The number of surviving cfu were normalized to cells that had not been treated with UV and represented as % survival. Strains containing plasmids pJV44 and pJV96 were empty vector control strains for pJV97 and pJV99 strains, respectively.

A.4 PRELIMINARY CONCLUSIONS

These experiments sought to assay for activity of the *M. smegmatis* RecBCD enzyme by two methods: (1) observing the UV phenotype of Δ*recB*, Δ*recD*, and Gam-expressing strains, and (2) by determining the recombineering frequencies in these strains. The UV sensitivity assay described above has been used previously in *M. smegmatis* with wild type and Δ*recA* strains, although the Δ*recA* strain survived ~100-fold better in this experiments than reported previously [155]. It is likely that this is due to variations in experimental procedures, since all repetitions of this assay consistently showed the same level of killing for both wild type and Δ*recA* strains. Additional assays could be tested, such as mitomycin C sensitivity, which may have more reproducible results. Further, a standard gene replacement of *M. smegmatis groEL1* was tested in each strain to determine the frequency of the recombination event compared to wild type.

The assays performed with λ Gam did not indicate that it had any effect on recombineering frequency or viability following UV treatment. It was not determined if the protein was expressed properly following induction, and it is also possible that the protein product was unstable or could not interact with the RecBCD complex. If protein expression and/or activity is an issue, alternative constructs could be tested. Another, possibly better, approach is to test additional proteins that inhibit nuclease activity that do not require protein-protein interactions, such as T4 gp2 or Mu Gam.

In *E. coli,* a Δ*recB* strain is decreased 30% in viability, ~100-fold in recombination activity, and 10-fold in survival in assays following treatment with DNA damaging agents [106,133]. These characteristics were not observed in the two *M. smegmatis* Δ*recB* strains; both behaved similarly to wild type with regard to overall viability and survival following UV

treatment. There are two likely, possibly independent, explanations for this. First, the two Δ*recB* strains – which were made separately – could have acquired suppressor mutations that restored UV-resistance, but retained the positive effect on recombineering frequency conferred by Δ*recB*. One category of suppressor mutations of Δ*recBC* in *E. coli* occurs in the *sbcB* and *sbcC* genes. These mutations restore wild type levels of recombination, DNA repair, and viability by inactivating the SbcCD and ExoI exonucleases. [104,118,221]. Suppressors of Δ*recBC* are also located in the *sbcA* gene only in certain strains of *E. coli*, which activate expression of *recET* (discussed in section 1.2.3). Therefore it is possible that the two *M. smegmatis* Δ*recB* strains used in these studies contained suppressor mutations that restored UV-resistance but were still deficient in dsDNA exonuclease activity such that recombineering frequencies were increased. It is not clear if mycobacteria, specifically *M. smegmatis*, encode *sbcBC*-like genes in which suppressor mutations could arise. It appears that *M. marium* encodes a protein with similarity to the SbcC of several *Bacillus* species, although it appears to be interrupted by a large (~475 amino acids) internal domain of unknown function. Further, the *M. marinum* SbcC does not have similarity to other mycobacterial proteins, and no SbcB or SbcC homologues have been identified thus far in other mycobacterial species.

An alternative possibility is that the AddA homologues – which are two different proteins for each mycobacterial genome – identified by bioinformatics form an active enzymatic complex that compensates for RecBCD. Specifically, the UV-resistant phenotype of the *M. smegmatis* Δ*recB* strain could be a result of AddA-like activity. In addition, there may be activities of the RecBCD complex that are not completely complemented by AddA, such as dsDNA exonuclease activity. This could explain why only a moderate increase in recombineering frequencies was observed in the Δ*recB* strain compared with the minimum of 10-fold increase observed in *E. coli*.

Additional experiments will be required to determine the role of the putative AddA proteins in the recombination and UV-sensitive phenotypes of *M. smegmatis* Δ*recB*.

Ultimately, the Δ*recB* strain showed a 3.7-fold increase in recombineering frequency, while the Δ*recD* strain improved frequencies to a lesser extent. However, the Δ*recB* strain could be useful for some recombineering approaches, particularly those in which phage genomes are the targets and the cell's genetic background is not of critical importance. In these assays, a relatively small number of phage DNA molecules are productively taken up by cells to produce plaques in a typical phage recombineering experiment (~100-200), and mutants are isolated at a frequency of 10-40%, or 1 out of 12-18 plaques. Therefore, increasing the frequency of recombination in the cells by using a Δ*recB* strain would enable easier identification of mutants for this particular application of recombineering, and potentially others.

BIBLIOGRAPHY

1. World Health Organization. (2008) *Global tuberculosis control report - surveillance, planning, financing 2008*. http://www.who.int/tb/publications/global_report/2008/en/index.html

2. Abraham, ZH and N Symonds (1990) Purification of overexpressed gam gene protein from bacteriophage Mu by denaturation-renaturation techniques and a study of its DNA-binding properties. *Biochem J* **269**(3): 679-84.

3. Aldovini, A, RN Husson and RA Young (1993) The uraA locus and homologous recombination in Mycobacterium bovis BCG. *J Bacteriol* **175**(22): 7282-9.

4. Altschul, SF, TL Madden, AA Schaffer, J Zhang, Z Zhang, W Miller and DJ Lipman (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**(17): 3389-402.

5. Amundsen, SK and GR Smith (2003) Interchangeable parts of the Escherichia coli recombination machinery. *Cell* **112**(6): 741-4.

6. Appasani, K, DS Thaler and EB Goldberg (1999) Bacteriophage T4 gp2 interferes with cell viability and with bacteriophage lambda Red recombination. *J Bacteriol* **181**(4): 1352-5.

7. Ausubel, FM, R Brent, RE Kingston, DD Moore, JG Siedman, JA Smith and K Struhl (2005) *Current protocols in molecular biology*. John Wiley & Sons, Inc.

8. Azad, AK, TD Sirakova, LM Rogers and PE Kolattukudy (1996) Targeted replacement of the mycocerosic acid synthase gene in Mycobacterium bovis BCG produces a mutant that lacks mycosides. *Proc Natl Acad Sci U S A* **93**(10): 4787-92.

9. Balasubramanian, V, MS Pavelka, Jr., SS Bardarov, J Martin, TR Weisbrod, RA McAdam, BR Bloom and WR Jacobs, Jr. (1996) Allelic exchange in Mycobacterium tuberculosis with long linear recombination substrates. *J Bacteriol* **178**(1): 273-9.

10. Banaiee, N, M Bobadilla-del-Valle, PF Riska, S Bardarov, Jr., PM Small, A Ponce-de-Leon, WR Jacobs, Jr., GF Hatfull and J Sifuentes-Osornio (2003) Rapid identification and susceptibility testing of Mycobacterium tuberculosis from MGIT cultures with luciferase reporter mycobacteriophages. *J Med Microbiol* **52**(Pt 7): 557-61.

11. Banerjee, A, E Dubnau, A Quemard, V Balasubramanian, KS Um, T Wilson, D Collins, G de Lisle and WR Jacobs, Jr. (1994) inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. *Science* **263**(5144): 227-30.

12. Barbour, SD, H Nagaishi, A Templin and AJ Clark (1970) Biochemical and genetic studies of recombination proficiency in Escherichia coli. II. Rec+ revertants caused by indirect suppression of rec- mutations. *Proc Natl Acad Sci U S A* **67**(1): 128-35.

13. Bardarov, S, J Kriakov, C Carriere, S Yu, C Vaamonde, RA McAdam, BR Bloom, GF Hatfull and WR Jacobs, Jr. (1997) Conditionally replicating mycobacteriophages: a system for transposon delivery to Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **94**(20): 10961-6.

14. Bardarov, S, S Bardarov Jr, Jr., MS Pavelka Jr, Jr., V Sambandamurthy, M Larsen, J Tufariello, J Chan, G Hatfull and WR Jacobs Jr, Jr. (2002) Specialized transduction: an efficient method for generating marked and unmarked targeted gene disruptions in Mycobacterium tuberculosis, M. bovis BCG and M. smegmatis. *Microbiology* **148**(Pt 10): 3007-17.

15. Baulard, A, L Kremer and C Locht (1996) Efficient homologous recombination in fast-growing and slow-growing mycobacteria. *J Bacteriol* **178**(11): 3091-8.

16. Baulard, A, L Kremer, P Supply, D Vidaud, JM Bidart, D Bellet and C Locht (1996) A new series of mycobacterial expression vectors for the development of live recombinant vaccines. *Gene* **176**(1-2): 149-54.

17. Bi, B, N Rybalchenko, EI Golub and CM Radding (2004) Human and yeast Rad52 proteins promote DNA strand exchange. *Proc Natl Acad Sci U S A* **101**(26): 9568-72.

18. Bibb, LA and GF Hatfull (2002) Integration and excision of the Mycobacterium tuberculosis prophage-like element, phiRv1. *Mol Microbiol* **45**(6): 1515-26.

19. Bloom, BR (1994) *Tuberculosis.* Washington, D.C.: ASM Press.

20. Bochman, ML and A Schwacha (2007) Differences in the single-stranded DNA binding activities of MCM2-7 and MCM467: MCM2 and MCM5 define a slow ATP-dependent step. *J Biol Chem* **282**(46): 33795-804.

21. Borsuk, S, TA Mendum, MQ Fagundes, M Michelon, CW Cunha, J McFadden and OA Dellagostin (2007) Auxotrophic complementation as a selectable marker for stable expression of foreign antigens in Mycobacterium bovis BCG. *Tuberculosis (Edinb)* **87**(6): 474-80.

22. Brodin, P, C Demangel and ST Cole (2005) Introduction to functional genomics of the *Mycobacterium tuberculosis* complex. In *Tuberculosis and the tubercle bacillus*. Vol. Eisenach, K, ST Cole, WR Jacobs, Jr. and D McMurray. Washington, D.C.: ASM Press, pp. 143-153.

23. Brooks, K and AJ Clark (1967) Behavior of lambda bacteriophage in a recombination deficienct strain of Escherichia coli. *J Virol* **1**(2): 283-93.

24. Brosch, R and MA Behr (2005) Comparative genomics and evolution of *Mycobacterium bovis* BCG. In *Tuberculosis and the tubercle bacillus*. Vol. Eisenach, K, ST Cole, WR Jacobs, Jr. and D McMurray. Washington, D.C.: ASM Press, pp. 155-164.

25. Brown, AC and T Parish (2006) Instability of the acetamide-inducible expression vector pJAM2 in Mycobacterium tuberculosis. *Plasmid* **55**(1): 81-6.

26. Calmette, A and C Guerin (1920) Nouvelles recherches experimentales sur la vaccination des bovides contre la tuberculose. *Ann Inst Pasteur* **34**: 553-560.

27. Carriere, C, PF Riska, O Zimhony, J Kriakov, S Bardarov, J Burns, J Chan and WR Jacobs, Jr. (1997) Conditionally replicating luciferase reporter phages: improved sensitivity for rapid detection and assessment of drug susceptibility of Mycobacterium tuberculosis. *J Clin Microbiol* **35**(12): 3232-9.

28. Cassuto, E, T Lash, KS Sriprakash and CM Radding (1971) Role of exonuclease and protein of phage lambda in genetic recombination. V. Recombination of lambda DNA in vitro. *Proc Natl Acad Sci U S A* **68**(7): 1639-43.

29. Cassuto, E and CM Radding (1971) Mechanism for the action of lambda exonuclease in genetic recombination. *Nat New Biol* **229**(1): 13-6.

30. Cha, RS, H Zarbl, P Keohavong and WG Thilly (1992) Mismatch amplification mutation assay (MAMA): application to the c-H-ras gene. *PCR Methods Appl* **2**(1): 14-20.

31. Chang, HW and DA Julin (2001) Structure and function of the Escherichia coli RecE protein, a member of the RecB nuclease domain family. *J Biol Chem* **276**(49): 46004-10.

32. Chedin, F and SC Kowalczykowski (2002) A novel family of regulated helicases/nucleases from Gram-positive bacteria: insights into the initiation of DNA recombination. *Mol Microbiol* **43**(4): 823-34.

33. Chu, CC, A Templin and AJ Clark (1989) Suppression of a frameshift mutation in the recE gene of Escherichia coli K-12 occurs by gene fusion. *J Bacteriol* **171**(4): 2101-9.

34. Clark, AJ (1974) Progress toward a metabolic interpretation of genetic recombination of Escherichia coli and bacteriophage lambda. *Genetics* **78**(1): 259-71.

35. Cole, ST, R Brosch, J Parkhill, T Garnier, C Churcher, D Harris, SV Gordon, K Eiglmeier, S Gas, CE Barry, 3rd, F Tekaia, K Badcock, D Basham, D Brown, T Chillingworth, R Connor, R Davies, K Devlin, T Feltwell, S Gentles, N Hamlin, S Holroyd, T Hornsby, K Jagels, A Krogh, J McLean, S Moule, L Murphy, K Oliver, J Osborne, MA Quail, MA Rajandream, J Rogers, S Rutter, K Seeger, J Skelton, R Squares, S Squares, JE Sulston, K Taylor, S Whitehead and BG Barrell (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393**(6685): 537-44.

36. Copeland, NG, NA Jenkins and DL Court (2001) Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet* **2**(10): 769-79.

37. Costantino, N and DL Court (2003) Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc Natl Acad Sci U S A* **100**(26): 15748-53.

38. Court, DL, JA Sawitzke and LC Thomason (2002) Genetic engineering using homologous recombination. *Annu Rev Genet* **36**: 361-88.

39. Court, R, N Cook, K Saikrishnan and D Wigley (2007) The crystal structure of lambda-Gam protein suggests a model for RecBCD inhibition. *J Mol Biol* **371**(1): 25-33.

40. Cuff, JA, ME Clamp, AS Siddiqui, M Finlay and GJ Barton (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**(10): 892-3.

41. Dahl, JL (2005) Scanning electron microscopy analysis of aged Mycobacterium tuberculosis cells. *Can J Microbiol* **51**(3): 277-81.

42. Datta, S, N Costantino and DL Court (2006) A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**: 109-15.

43. Datta, S, N Costantino, X Zhou and DL Court (2008) Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc Natl Acad Sci U S A* **105**(5): 1626-31.

44. Daugelat, S, J Kowall, J Mattow, D Bumann, R Winter, R Hurwitz and SH Kaufmann (2003) The RD1 proteins of Mycobacterium tuberculosis: expression in Mycobacterium smegmatis and biochemical characterization. *Microbes Infect* **5**(12): 1082-95.

45. Davis, EO, SG Sedgwick and MJ Colston (1991) Novel structure of the recA locus of Mycobacterium tuberculosis implies processing of the gene product. *J Bacteriol* **173**(18): 5653-62.

46. Davis, EO, PJ Jenner, PC Brooks, MJ Colston and SG Sedgwick (1992) Protein splicing in the maturation of M. tuberculosis recA protein: a mechanism for tolerating a novel class of intervening sequence. *Cell* **71**(2): 201-10.

47. Davis, EO, HS Thangaraj, PC Brooks and MJ Colston (1994) Evidence of selection for protein introns in the recAs of pathogenic mycobacteria. *Embo J* **13**(3): 699-703.

48. Davis, EO, B Springer, KK Gopaul, KG Papavinasasundaram, P Sander and EC Bottger (2002) DNA damage induction of recA in Mycobacterium tuberculosis independently of RecA and LexA. *Mol Microbiol* **46**(3): 791-800.

49. Donnelly-Wu, MK, WR Jacobs, Jr. and GF Hatfull (1993) Superinfection immunity of mycobacteriophage L5: applications for genetic transformation of mycobacteria. *Mol Microbiol* **7**(3): 407-17.

50. Ehrlich, SD, H Bierne, E d'Alencon, D Vilette, M Petranovic, P Noirot and B Michel (1993) Mechanisms of illegitimate recombination. *Gene* **135**(1-2): 161-6.

51. Ehrt, S, XV Guo, CM Hickey, M Ryou, M Monteleone, LW Riley and D Schnappinger (2005) Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Res* **33**(2): e21.

52. Ellis, HM, D Yu, T DiTizio and DL Court (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci U S A* **98**(12): 6742-6.

53. Enquist, LW and A Skalka (1973) Replication of bacteriophage lambda DNA dependent on the function of host and viral genes. I. Interaction of red, gam and rec. *J Mol Biol* **75**(2): 185-212.

54. Fenton, AC and AR Poteete (1984) Genetic analysis of the erf region of the bacteriophage P22 chromosome. *Virology* **134**(1): 148-60.

55. Fishel, RA, AA James and R Kolodner (1981) recA-independent general genetic recombination of plasmids. *Nature* **294**(5837): 184-6.

56. Frischkorn, K, P Sander, M Scholz, K Teschner, T Prammananan and EC Bottger (1998) Investigation of mycobacterial recA function: protein introns in the RecA of pathogenic

mycobacteria do not affect competency for homologous recombination. *Mol Microbiol* **29**(5): 1203-14.

57. Frischkorn, K, B Springer, EC Bottger, EO Davis, MJ Colston and P Sander (2000) In vivo splicing and functional characterization of Mycobacterium leprae RecA. *J Bacteriol* **182**(12): 3590-2.

58. Furth, ME and SH Wickner (1983) Lambda DNA replication. In *Lambda II*. Vol. Hendrix, RW, JW Roberts, FW Stahl and RA Wiesberg. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, pp. 145-174.

59. Garbe, TR, J Barathi, S Barnini, Y Zhang, C Abou-Zeid, D Tang, R Mukherjee and DB Young (1994) Transformation of mycobacterial species using hygromycin resistance as selectable marker. *Microbiology* **140 (Pt 1)**: 133-8.

60. Ghosh, P, LR Wasil and GF Hatfull (2006) Control of Phage Bxb1 Excision by a Novel Recombination Directionality Factor. *PLoS Biol* **4**(6): e186.

61. Ghosh, P, LA Bibb and GF Hatfull (2008) Two-step site selection for serine-integrase-mediated excision: DNA-directed integrase conformation and central dinucleotide proofreading. *Proc Natl Acad Sci U S A* **105**(9): 3238-43.

62. Gicquel-Sanzey, B, J Moniz-Pereira, M Gheorghiu and J Rauzier (1989) Structure of pAL5000, a plasmid from M. fortuitum and its utilization in transformation of mycobacteria. *Acta Leprol* **7 Suppl 1**: 208-11.

63. Gillen, JR, AE Karu, H Nagaishi and AJ Clark (1977) Characterization of the deoxyribonuclease determined by lambda reverse as exonuclease VIII of Escherichia coli. *J Mol Biol* **113**(1): 27-41.

64. Gillen, JR, DK Willis and AJ Clark (1981) Genetic analysis of the RecE pathway of genetic recombination in Escherichia coli K-12. *J Bacteriol* **145**(1): 521-32.

65. Gopaul, KK, PC Brooks, JF Prost and EO Davis (2003) Characterization of the two Mycobacterium tuberculosis recA promoters. *J Bacteriol* **185**(20): 6005-15.

66. Gottesman, MM, ME Gottesman, S Gottesman and M Gellert (1974) Characterization of bacteriophage lambda reverse as an Escherichia coli phage carrying a unique set of host-derived recombination functions. *J Mol Biol* **88**(2): 471-87.

67. Guo, XV, M Monteleone, M Klotzsche, A Kamionka, W Hillen, M Braunstein, S Ehrt and D Schnappinger (2007) Silencing Mycobacterium smegmatis by using tetracycline repressors. *J Bacteriol* **189**(13): 4614-23.

68. Hall, SD, MF Kane and RD Kolodner (1993) Identification and characterization of the Escherichia coli RecT protein, a protein encoded by the recE region that promotes renaturation of homologous single-stranded DNA. *J Bacteriol* **175**(1): 277-87.

69. Hall, SD and RD Kolodner (1994) Homologous pairing and strand exchange promoted by the Escherichia coli RecT protein. *Proc Natl Acad Sci U S A* **91**(8): 3205-9.

70. Hatfull, GF, L Barsom, L Chang, M Donnelly-Wu, MH Lee, M Levin, C Nesbit and GJ Sarkis (1994) Bacteriophages as tools for vaccine development. *Dev Biol Stand* **82**: 43-7.

71. Hatfull, GF (2004) Mycobacteriophages and tuberculosis. In *Tuberculosis*. Vol. Eisenach, K, ST Cole, WR Jacobs, Jr. and D McMurray. Washington, D.C.: ASM Press, pp. 203-218.

72. Hatfull, GF (2005) Mycobacteriophages: pathogenesis and applications. In *Phages: their role in bacterial pathogenesis and biotechnology*. Vol. Waldor, MK, DI Friedman and SL Adhya. Washington, D.C.: ASM Press, pp. 238-255.

73. Hatfull, GF, ML Pedulla, D Jacobs-Sera, PM Cichon, A Foley, ME Ford, RM Gonda, JM Houtz, AJ Hryckowian, VA Kelchner, S Namburi, KV Pajcini, MG Popovich, DT Schleicher, BZ Simanek, AL Smith, GM Zdanowicz, V Kumar, CL Peebles, WR Jacobs, Jr., JG Lawrence and RW Hendrix (2006) Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. *PLoS Genet* **2**(6): e92.

74. Hendrickson, H and JG Lawrence (2006) Selection for chromosome architecture in bacteria. *J Mol Evol* **62**(5): 615-29.

75. Hendrickson, H and JG Lawrence (2007) Mutational bias suggests that replication termination occurs near the dif site, not at Ter sites. *Mol Microbiol* **64**(1): 42-56.

76. Hendrix, RW, MC Smith, RN Burns, ME Ford and GF Hatfull (1999) Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci U S A* **96**(5): 2192-7.

77. Hendrix, RW (2002) Bacteriophages: evolution of the majority. *Theor Popul Biol* **61**(4): 471-80.

78. Hermans, J, C Martin, GN Huijberts, T Goosen and JA de Bont (1991) Transformation of Mycobacterium aurum and Mycobacterium smegmatis with the broad host-range gram-negative cosmid vector pJRD215. *Mol Microbiol* **5**(6): 1561-6.

79. Hill, SA, MM Stahl and FW Stahl (1997) Single-strand DNA intermediates in phage lambda's Red recombination pathway. *Proc Natl Acad Sci U S A* **94**(7): 2951-6.

80. Hinds, J, E Mahenthiralingam, KE Kempsell, K Duncan, RW Stokes, T Parish and NG Stoker (1999) Enhanced gene replacement in mycobacteria. *Microbiology* **145 (Pt 3)**: 519-27.

81. Hondalus, MK, S Bardarov, R Russell, J Chan, WR Jacobs, Jr. and BR Bloom (2000) Attenuation of and protection induced by a leucine auxotroph of Mycobacterium tuberculosis. *Infect Immun* **68**(5): 2888-98.

82. Husson, RN, BE James and RA Young (1990) Gene replacement and expression of foreign DNA in mycobacteria. *J Bacteriol* **172**(2): 519-24.

83. Igoucheva, O, V Alexeev and K Yoon (2001) Targeted gene correction by small single-stranded oligonucleotides in mammalian cells. *Gene Ther* **8**(5): 391-9.

84. Ikeda, H, K Shiraishi and Y Ogata (2004) Illegitimate recombination mediated by double-strand break and end-joining in Escherichia coli. *Adv Biophys* **38**: 3-20.

85. Iyer, LM, EV Koonin and L Aravind (2002) Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC Genomics* **3**(1): 8.

86. Jacobs, WR, Jr., M Tuckman and BR Bloom (1987) Introduction of foreign DNA into mycobacteria using a shuttle phasmid. *Nature* **327**(6122): 532-5.

87. Jacobs, WR, Jr., SB Snapper, L Lugosi, A Jekkel, RE Melton, T Kieser and BR Bloom (1989) Development of genetic systems for the mycobacteria. *Acta Leprol* **7 Suppl 1**: 203-7.

88. Jacobs, WR, Jr., RG Barletta, R Udani, J Chan, G Kalkut, G Sosne, T Kieser, GJ Sarkis, GF Hatfull and BR Bloom (1993) Rapid assessment of drug susceptibilities of Mycobacterium tuberculosis by means of luciferase reporter phages. *Science* **260**(5109): 819-22.

89. Joseph, JW and R Kolodner (1983) Exonuclease VIII of Escherichia coli. II. Mechanism of action. *J Biol Chem* **258**(17): 10418-24.

90. Joseph, JW and R Kolodner (1983) Exonuclease VIII of Escherichia coli. I. Purification and physical properties. *J Biol Chem* **258**(17): 10411-7.

91. Kalpana, GV, BR Bloom and WR Jacobs, Jr. (1991) Insertional mutagenesis and illegitimate recombination in mycobacteria. *Proc Natl Acad Sci U S A* **88**(12): 5433-7.

92. Karakousis, G, N Ye, Z Li, SK Chiu, G Reddy and CM Radding (1998) The beta protein of phage lambda binds preferentially to an intermediate in DNA renaturation. *J Mol Biol* **276**(4): 721-31.

93. Karu, AE, Y Sakaki, H Echols and S Linn (1975) The gamma protein specified by bacteriophage gamma. Structure and inhibitory activity for the recBC enzyme of Escherichia coli. *J Biol Chem* **250**(18): 7377-87.

94. Karunakaran, P and J Davies (2000) Genetic antagonism and hypermutability in Mycobacterium smegmatis. *J Bacteriol* **182**(12): 3331-5.

95. Kim, AI, P Ghosh, MA Aaron, LA Bibb, S Jain and GF Hatfull (2003) Mycobacteriophage Bxb1 integrates into the Mycobacterium smegmatis groEL1 gene. *Mol Microbiol* **50**(2): 463-73.

96. Kmiec, E and WK Holloman (1981) Beta protein of bacteriophage lambda promotes renaturation of DNA. *J Biol Chem* **256**(24): 12636-9.

97. Knipfer, N, A Seth and TE Shrader (1997) Unmarked gene integration into the chromosome of Mycobacterium smegmatis via precise replacement of the pyrF gene. *Plasmid* **37**(2): 129-40.

98. Kooistra, J, BJ Haijema and G Venema (1993) The Bacillus subtilis addAB genes are fully functional in Escherichia coli. *Mol Microbiol* **7**(6): 915-23.

99. Kornberg, A and TA Baker (1992) *DNA Replication.* New York: W. H. Freeman and Company.

100. Kovall, R and BW Matthews (1997) Toroidal structure of lambda-exonuclease. *Science* **277**(5333): 1824-7.

101. Krzysiak, TC, T Wendt, LR Sproul, P Tittmann, H Gross, SP Gilbert and A Hoenger (2006) A structural model for monastrol inhibition of dimeric kinesin Eg5. *Embo J* **25**(10): 2263-73.

102. Kulkarni, SK and FW Stahl (1989) Interaction between the sbcC gene of Escherichia coli and the gam gene of phage lambda. *Genetics* **123**(2): 249-53.

103. Kumar, RA, MB Vaze, NR Chandra, M Vijayan and K Muniyappa (1996) Functional characterization of the precursor and spliced forms of RecA protein of Mycobacterium tuberculosis. *Biochemistry* **35**(6): 1793-802.

104. Kushner, SR, H Nagaishi, A Templin and AJ Clark (1971) Genetic recombination in Escherichia coli: the role of exonuclease I. *Proc Natl Acad Sci U S A* **68**(4): 824-7.

105. Kushner, SR, H Nagaishi and AJ Clark (1974) Isolation of exonuclease VIII: the enzyme associated with sbcA indirect suppressor. *Proc Natl Acad Sci U S A* **71**(9): 3593-7.

106. Kuzminov, A (1999) Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda. *Microbiol Mol Biol Rev* **63**(4): 751-813, table of contents.

107. Labidi, A, HL David and D Roulland-Dussoix (1985) Restriction endonuclease mapping and cloning of Mycobacterium fortuitum var. fortuitum plasmid pAL5000. *Ann Inst Pasteur Microbiol* **136B**(2): 209-15.

108. Lawrence, JG, GF Hatfull and RW Hendrix (2002) Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. *J Bacteriol* **184**(17): 4891-905.

109. Lee, EC, D Yu, J Martinez de Velasco, L Tessarollo, DA Swing, DL Court, NA Jenkins and NG Copeland (2001) A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**(1): 56-65.

110. Lee, MH, L Pascopella, WR Jacobs, Jr. and GF Hatfull (1991) Site-specific integration of mycobacteriophage L5: integration-proficient vectors for Mycobacterium smegmatis, Mycobacterium tuberculosis, and bacille Calmette-Guerin. *Proc Natl Acad Sci U S A* **88**(8): 3111-5.

111. Lee, S, J Kriakov, C Vilcheze, Z Dai, GF Hatfull and WR Jacobs, Jr. (2004) Bxz1, a new generalized transducing phage for mycobacteria. *FEMS Microbiol Lett* **241**(2): 271-6.

112. Lety, MA, S Nair, P Berche and V Escuyer (1997) A single point mutation in the embB gene is responsible for resistance to ethambutol in Mycobacterium smegmatis. *Antimicrob Agents Chemother* **41**(12): 2629-33.

113. Li, XT, N Costantino, LY Lu, DP Liu, RM Watt, KS Cheah, DL Court and JD Huang (2003) Identification of factors influencing strand bias in oligonucleotide-mediated recombination in Escherichia coli. *Nucleic Acids Res* **31**(22): 6674-87.

114. Li, Z, G Karakousis, SK Chiu, G Reddy and CM Radding (1998) The beta protein of phage lambda promotes strand exchange. *J Mol Biol* **276**(4): 733-44.

115. Lipinska, B, AS Rao, BM Bolten, R Balakrishnan and EB Goldberg (1989) Cloning and identification of bacteriophage T4 gene 2 product gp2 and action of gp2 on infecting DNA in vivo. *J Bacteriol* **171**(1): 488-97.

116. Little, JW (1967) An exonuclease induced by bacteriophage lambda. II. Nature of the enzymatic reaction. *J Biol Chem* **242**(4): 679-86.

117. Liu, L, MC Rice, M Drury, S Cheng, H Gamper and EB Kmiec (2002) Strand bias in targeted gene repair is influenced by transcriptional activity. *Mol Cell Biol* **22**(11): 3852-63.

118. Lloyd, RG and C Buckman (1985) Identification and genetic analysis of sbcC mutations in commonly used recBC sbcB strains of Escherichia coli K-12. *J Bacteriol* **164**(2): 836-44.

119. Luisi-DeLuca, C, AJ Clark and RD Kolodner (1988) Analysis of the recE locus of Escherichia coli K-12 by use of polyclonal antibodies to exonuclease VIII. *J Bacteriol* **170**(12): 5797-805.

120. Malyarchuk, S, D Wright, R Castore, E Klepper, B Weiss, AJ Doherty and L Harrison (2007) Expression of Mycobacterium tuberculosis Ku and Ligase D in Escherichia coli results in RecA and RecB-independent DNA end-joining at regions of microhomology. *DNA Repair (Amst)* **6**(10): 1413-24.

121. Marklund, BI, DP Speert and RW Stokes (1995) Gene replacement through homologous recombination in Mycobacterium intracellulare. *J Bacteriol* **177**(21): 6100-5.

122. Marsic, N, S Roje, I Stojiljkovic, E Salaj-Smic and Z Trgovcevic (1993) In vivo studies on the interaction of RecBCD enzyme and lambda Gam protein. *J Bacteriol* **175**(15): 4738-43.

123. Martinsohn, JT, M Radman and MA Petit (2008) The lambda red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism. *PLoS Genet* **4**(5): e1000065.

124. Matsuura, S, J Komatsu, K Hirano, H Yasuda, K Takashima, S Katsura and A Mizuno (2001) Real-time observation of a single DNA digestion by lambda exonuclease under a fluorescence microscope field. *Nucleic Acids Res* **29**(16): E79.

125. McFadden, J (1996) Recombination in mycobacteria. *Mol Microbiol* **21**(2): 205-11.

126. Morris, P, LJ Marinelli, D Jacobs-Sera, RW Hendrix and GF Hatfull (2008) Genomic characterization of mycobacteriophage giles: evidence for phage acquisition of host DNA by illegitimate recombination. *J Bacteriol* **190**(6): 2172-82.

127. Movahedzadeh, F, MJ Colston and EO Davis (1997) Determination of DNA sequences required for regulated Mycobacterium tuberculosis RecA expression in response to DNA-damaging agents suggests that two modes of regulation exist. *J Bacteriol* **179**(11): 3509-18.

128. Movahedzadeh, F, MJ Colston and EO Davis (1997) Characterization of Mycobacterium tuberculosis LexA: recognition of a Cheo (Bacillus-type SOS) box. *Microbiology* **143 (Pt 3)**: 929-36.

129. Muniyappa, K and CM Radding (1986) The homologous recombination system of phage lambda. Pairing activities of beta protein. *J Biol Chem* **261**(16): 7472-8.

130. Muniyappa, K, MB Vaze, N Ganesh, M Sreedhar Reddy, N Guhan and R Venkatesh (2000) Comparative genomics of Mycobacterium tuberculosis and Escherichia coli for recombination (rec) genes. *Microbiology* **146 (Pt 9)**: 2093-5.

131. Murphy, KC, L Casey, N Yannoutsos, AR Poteete and RW Hendrix (1987) Localization of a DNA-binding determinant in the bacteriophage P22 Erf protein. *J Mol Biol* **194**(1): 105-17.

132. Murphy, KC, AC Fenton and AR Poteete (1987) Sequence of the bacteriophage P22 anti-recBCD (abc) genes and properties of P22 abc region deletion mutants. *Virology* **160**(2): 456-64.

133. Murphy, KC (1991) Lambda Gam protein inhibits the helicase and chi-stimulated recombination activities of Escherichia coli RecBCD enzyme. *J Bacteriol* **173**(18): 5808-21.

134. Murphy, KC (1994) Biochemical characterization of P22 phage-modified Escherichia coli RecBCD enzyme. *J Biol Chem* **269**(36): 22507-16.

135. Murphy, KC (1998) Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. *J Bacteriol* **180**(8): 2063-71.

136. Murphy, KC (2000) Bacteriophage P22 Abc2 protein binds to RecC increases the 5' strand nicking activity of RecBCD and together with lambda bet, promotes Chi-independent recombination. *J Mol Biol* **296**(2): 385-401.

137. Murphy, KC, KG Campellone and AR Poteete (2000) PCR-mediated gene replacement in Escherichia coli. *Gene* **246**(1-2): 321-30.

138. Murphy, KC (2007) The lambda Gam protein inhibits RecBCD binding to dsDNA ends. *J Mol Biol* **371**(1): 19-24.

139. Muttucumaru, DG and T Parish (2004) The molecular biology of recombination in Mycobacteria: what do we know and how can we use it? *Curr Issues Mol Biol* **6**(2): 145-57.

140. Muyrers, JP, Y Zhang, G Testa and AF Stewart (1999) Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res* **27**(6): 1555-7.

141. Muyrers, JP, Y Zhang, V Benes, G Testa, W Ansorge and AF Stewart (2000) Point mutation of bacterial artificial chromosomes by ET recombination. *EMBO Rep* **1**(3): 239-43.

142. Muyrers, JP, Y Zhang, F Buchholz and AF Stewart (2000) RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev* **14**(15): 1971-82.

143. Muyrers, JP, Y Zhang and AF Stewart (2000) ET-cloning: think recombination first. *Genet Eng (N Y)* **22**: 77-98.

144. Mythili, E, KA Kumar and K Muniyappa (1996) Characterization of the DNA-binding domain of beta protein, a component of phage lambda red-pathway, by UV catalyzed cross-linking. *Gene* **182**(1-2): 81-7.

145. Noirot, P and RD Kolodner (1998) DNA strand invasion promoted by Escherichia coli RecT protein. *J Biol Chem* **273**(20): 12274-80.

146. Noirot, P, RC Gupta, CM Radding and RD Kolodner (2003) Hallmarks of homology recognition by RecA-like recombinases are exhibited by the unrelated Escherichia coli RecT protein. *Embo J* **22**(2): 324-34.

147. Norman, E, OA Dellagostin, J McFadden and JW Dale (1995) Gene replacement by homologous recombination in Mycobacterium bovis BCG. *Mol Microbiol* **16**(4): 755-60.

148. Notredame, C, DG Higgins and J Heringa (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**(1): 205-17.

149. Ojha, A, M Anand, A Bhatt, L Kremer, WR Jacobs, Jr. and GF Hatfull (2005) GroEL1: A Dedicated Chaperone Involved in Mycolic Acid Biosynthesis during Biofilm Formation in Mycobacteria. *Cell* **123**(5): 861-73.

150. Ojha, AK, AD Baughn, D Sambandan, T Hsu, X Trivelli, Y Guerardel, A Alahari, L Kremer, WR Jacobs, Jr. and GF Hatfull (2008) Growth of Mycobacterium tuberculosis biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. *Mol Microbiol*.

151. Oppenheim, AB, AJ Rattray, M Bubunenko, LC Thomason and DL Court (2004) In vivo recombineering of bacteriophage lambda by PCR fragments and single-strand oligonucleotides. *Virology* **319**(2): 185-9.

152. Orr-Weaver, TL, JW Szostak and RJ Rothstein (1981) Yeast transformation: a model system for the study of recombination. *Proc Natl Acad Sci U S A* **78**(10): 6354-8.

153. Paget, E and J Davies (1996) Apramycin resistance as a selective marker for gene transfer in mycobacteria. *J Bacteriol* **178**(21): 6357-60.

154. Papavinasasundaram, KG, F Movahedzadeh, JT Keer, NG Stoker, MJ Colston and EO Davis (1997) Mycobacterial recA is cotranscribed with a potential regulatory gene called recX. *Mol Microbiol* **24**(1): 141-53.

155. Papavinasasundaram, KG, MJ Colston and EO Davis (1998) Construction and complementation of a recA deletion mutant of Mycobacterium smegmatis reveals that the intein in Mycobacterium tuberculosis recA does not affect RecA function. *Mol Microbiol* **30**(3): 525-34.

156. Papavinasasundaram, KG, C Anderson, PC Brooks, NA Thomas, F Movahedzadeh, PJ Jenner, MJ Colston and EO Davis (2001) Slow induction of RecA by DNA damage in Mycobacterium tuberculosis. *Microbiology* **147**(Pt 12): 3271-9.

157. Parish, T, E Mahenthiralingam, P Draper, EO Davis and MJ Colston (1997) Regulation of the inducible acetamidase gene of Mycobacterium smegmatis. *Microbiology* **143 (Pt 7)**: 2267-76.

158. Parish, T, BG Gordhan, RA McAdam, K Duncan, V Mizrahi and NG Stoker (1999) Production of mutants in amino acid biosynthesis genes of Mycobacterium tuberculosis by homologous recombination. *Microbiology* **145 (Pt 12)**: 3497-503.

159. Parish, T and NG Stoker (2000) Use of a flexible cassette method to generate a double unmarked Mycobacterium tuberculosis tlyA plcABC mutant by gene replacement. *Microbiology* **146 (Pt 8)**: 1969-75.

160. Pashley, C and NG Stoker (2000) Plasmids in mycobacteria. In *Molecular genetics of mycobacteria*. Vol. Hatfull, GF and WR Jacobs, Jr. Washington, D.C.: ASM Press, pp. 55-68.

161. Pashley, CA, T Parish, RA McAdam, K Duncan and NG Stoker (2003) Gene replacement in mycobacteria by using incompatible plasmids. *Appl Environ Microbiol* **69**(1): 517-23.

162. Passy, SI, X Yu, Z Li, CM Radding and EH Egelman (1999) Rings and filaments of beta protein from bacteriophage lambda suggest a superfamily of recombination proteins. *Proc Natl Acad Sci U S A* **96**(8): 4279-84.

163. Pavelka, MS, Jr. and WR Jacobs, Jr. (1999) Comparison of the construction of unmarked deletion mutations in Mycobacterium smegmatis, Mycobacterium bovis bacillus Calmette-Guerin, and Mycobacterium tuberculosis H37Rv by allelic exchange. *J Bacteriol* **181**(16): 4780-9.

164. Pearson, RE, S Jurgensen, GJ Sarkis, GF Hatfull and WR Jacobs, Jr. (1996) Construction of D29 shuttle phasmids and luciferase reporter phages for detection of mycobacteria. *Gene* **183**(1-2): 129-36.

165. Pedulla, ML, ME Ford, JM Houtz, T Karthikeyan, C Wadsworth, JA Lewis, D Jacobs-Sera, J Falbo, J Gross, NR Pannunzio, W Brucker, V Kumar, J Kandasamy, L Keenan, S Bardarov, J Kriakov, JG Lawrence, WR Jacobs, Jr., RW Hendrix and GF Hatfull (2003) Origins of highly mosaic mycobacteriophage genomes. *Cell* **113**(2): 171-82.

166. Pelicic, V, JM Reyrat and B Gicquel (1996) Expression of the Bacillus subtilis sacB gene confers sucrose sensitivity on mycobacteria. *J Bacteriol* **178**(4): 1197-9.

167. Pelicic, V, JM Reyrat and B Gicquel (1996) Positive selection of allelic exchange mutants in Mycobacterium bovis BCG. *FEMS Microbiol Lett* **144**(2-3): 161-6.

168. Pelicic, V, JM Reyrat and B Gicquel (1996) Generation of unmarked directed mutations in mycobacteria, using sucrose counter-selectable suicide vectors. *Mol Microbiol* **20**(5): 919-25.

169. Pelicic, V, M Jackson, JM Reyrat, WR Jacobs, Jr., B Gicquel and C Guilhot (1997) Efficient allelic exchange and transposon mutagenesis in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **94**(20): 10955-60.

170. Pham, TT, D Jacobs-Sera, ML Pedulla, RW Hendrix and GF Hatfull (2007) Comparative genomic analysis of mycobacteriophage Tweety: evolutionary insights and construction of compatible site-specific integration vectors for mycobacteria. *Microbiology* **153**(Pt 8): 2711-23.

171. Pierce, EA, Q Liu, O Igoucheva, R Omarrudin, H Ma, SL Diamond and K Yoon (2003) Oligonucleotide-directed single-base DNA alterations in mouse embryonic stem cells. *Gene Ther* **10**(1): 24-33.

172. Piuri, M and GF Hatfull (2006) A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells. *Mol Microbiol* **62**(6): 1569-85.

173. Poteete, AR and AC Fenton (1983) DNA-binding properties of the Erf protein of bacteriophage P22. *J Mol Biol* **163**(2): 257-75.

174. Poteete, AR, RT Sauer and RW Hendrix (1983) Domain structure and quaternary organization of the bacteriophage P22 Erf protein. *J Mol Biol* **171**(4): 401-18.

175. Poteete, AR and AC Fenton (1984) Lambda red-dependent growth and recombination of phage P22. *Virology* **134**(1): 161-7.

176. Poteete, AR, AC Fenton and KC Murphy (1988) Modulation of Escherichia coli RecBCD activity by the bacteriophage lambda Gam and P22 Abc functions. *J Bacteriol* **170**(5): 2012-21.

177. Poteete, AR, AC Fenton and AV Semerjian (1991) Bacteriophage P22 accessory recombination function. *Virology* **182**(1): 316-23.

178. Poteete, AR and AC Fenton (1993) Efficient double-strand break-stimulated recombination promoted by the general recombination systems of phages lambda and P22. *Genetics* **134**(4): 1013-21.

179. Poteete, AR (2001) What makes the bacteriophage lambda Red system useful for genetic engineering: molecular mechanism and biological function. *FEMS Microbiol Lett* **201**(1): 9-14.

180. Poteete, AR (2008) Involvement of DNA replication in phage lambda Red-mediated homologous recombination. *Mol Microbiol* **68**(1): 66-74.

181. Radding, CM (1970) The role of exonuclease and beta protein of bacteriophage lambda in genetic recombination. I. Effects of red mutants on protein structure. *J Mol Biol* **52**(3): 491-9.

182. Radding, CM and DM Carter (1971) The role of exonuclease and beta protein of phage lambda in genetic recombination. 3. Binding to deoxyribonucleic acid. *J Biol Chem* **246**(8): 2513-8.

183. Raj, CV and T Ramakrishnan (1970) Transduction in Mycobacterium smegmatis. *Nature* **228**(5268): 280-1.

184. Ramakrishnan, L, HT Tran, NA Federspiel and S Falkow (1997) A crtB homolog essential for photochromogenicity in Mycobacterium marinum: isolation, characterization, and gene disruption via homologous recombination. *J Bacteriol* **179**(18): 5862-8.

185. Ranallo, RT, S Barnoy, S Thakkar, T Urick and MM Venkatesan (2006) Developing live Shigella vaccines using lambda Red recombineering. *FEMS Immunol Med Microbiol* **47**(3): 462-9.

186. Rand, L, J Hinds, B Springer, P Sander, RS Buxton and EO Davis (2003) The majority of inducible DNA repair genes in Mycobacterium tuberculosis are induced independently of RecA. *Mol Microbiol* **50**(3): 1031-42.

187. Revel, V, E Cambau, V Jarlier and W Sougakoff (1994) Characterization of mutations in Mycobacterium smegmatis involved in resistance to fluoroquinolones. *Antimicrob Agents Chemother* **38**(9): 1991-6.

188. Reyrat, JM, FX Berthet and B Gicquel (1995) The urease locus of Mycobacterium tuberculosis and its utilization for the demonstration of allelic exchange in Mycobacterium bovis bacillus Calmette-Guerin. *Proc Natl Acad Sci U S A* **92**(19): 8768-72.

189. Riska, PF, Y Su, S Bardarov, L Freundlich, G Sarkis, G Hatfull, C Carriere, V Kumar, J Chan and WR Jacobs, Jr. (1999) Rapid film-based determination of antibiotic susceptibilities of Mycobacterium tuberculosis strains by using a luciferase reporter phage and the Bronx Box. *J Clin Microbiol* **37**(4): 1144-9.

190. Roberts, G, DG Muttucumaru and T Parish (2003) Control of the acetamidase gene of Mycobacterium smegmatis by multiple regulators. *FEMS Microbiol Lett* **221**(1): 131-6.

191. Rouse, DA, JA DeVito, Z Li, H Byer and SL Morris (1996) Site-directed mutagenesis of the katG gene of Mycobacterium tuberculosis: effects on catalase-peroxidase activities and isoniazid resistance. *Mol Microbiol* **22**(3): 583-92.

192. Russell, CB, DS Thaler and FW Dahlquist (1989) Chromosomal transformation of Escherichia coli recD strains with linearized plasmids. *J Bacteriol* **171**(5): 2609-13.

193. Rybalchenko, N, EI Golub, B Bi and CM Radding (2004) Strand invasion promoted by recombination protein beta of coliphage lambda. *Proc Natl Acad Sci U S A* **101**(49): 17056-60.

194. Sacchettini, JC, EJ Rubin and JS Freundlich (2008) Drugs versus bugs: in pursuit of the persistent predator Mycobacterium tuberculosis. *Nat Rev Microbiol* **6**(1): 41-52.

195. Sakaki, Y (1974) Inactivation of the ATP-dependent DNase of Escherichia coli after infection with double-stranded DNA phages. *J Virol* **14**(6): 1611-2.

196. Sander, P, A Meier and EC Bottger (1995) rpsL+: a dominant selectable marker for gene replacement in mycobacteria. *Mol Microbiol* **16**(5): 991-1000.

197. Sarkis, GJ, WR Jacobs, Jr. and GF Hatfull (1995) L5 luciferase reporter mycobacteriophages: a sensitive tool for the detection and assay of live mycobacteria. *Mol Microbiol* **15**(6): 1055-67.

198. Sarkis, GJ and GF Hatfull (1998) Mycobacteriophages. *Methods Mol Biol* **101**: 145-73.

199. Sarov, M, S Schneider, A Pozniakovski, A Roguev, S Ernst, Y Zhang, AA Hyman and AF Stewart (2006) A recombineering pipeline for functional genomics applied to Caenorhabditis elegans. *Nat Methods* **3**(10): 839-44.

200. Sassetti, CM, DH Boyd and EJ Rubin (2001) Comprehensive identification of conditionally essential genes in mycobacteria. *Proc Natl Acad Sci U S A* **98**(22): 12712-7.

201. Sawitzke, JA, LC Thomason, N Costantino, M Bubunenko, S Datta and DL Court (2007) Recombineering: in vivo genetic engineering in E. coli, S. enterica, and beyond. *Methods Enzymol* **421**: 171-99.

202. Scollard, DM, LB Adams, TP Gillis, JL Krahenbuhl, RW Truman and DL Williams (2006) The continuing challenges of leprosy. *Clin Microbiol Rev* **19**(2): 338-81.

203. Semerjian, AV, DC Malloy and AR Poteete (1989) Genetic structure of the bacteriophage P22 PL operon. *J Mol Biol* **207**(1): 1-13.

204. Shinohara, A, M Shinohara, T Ohta, S Matsuda and T Ogawa (1998) Rad52 forms ring structures and co-operates with RPA in single-strand DNA annealing. *Genes Cells* **3**(3): 145-56.

205. Shiraishi, K, K Hanada, Y Iwakura and H Ikeda (2002) Roles of RecJ, RecO, and RecR in RecET-mediated illegitimate recombination in Escherichia coli. *J Bacteriol* **184**(17): 4715-21.

206. Shulman, MJ, LM Hallick, H Echols and ER Signer (1970) Properties of recombination-deficient mutants of bacteriophage lambda. *J Mol Biol* **52**(3): 501-20.

207. Signer, ER and J Weil (1968) Recombination in bacteriophage lambda. I. Mutants deficient in general recombination. *J Mol Biol* **34**(2): 261-71.

208. Silverstein, JL and EB Goldberg (1976) T4 DNA injection. II. Protection of entering DNA from host exonuclease V. *Virology* **72**(1): 212-23.

209. Slayden, RA and CE Barry, 3rd (2000) The genetics and biochemistry of isoniazid resistance in mycobacterium tuberculosis. *Microbes Infect* **2**(6): 659-69.

210. Snapper, SB, L Lugosi, A Jekkel, RE Melton, T Kieser, BR Bloom and WR Jacobs, Jr. (1988) Lysogeny and transformation in mycobacteria: stable expression of foreign genes. *Proc Natl Acad Sci U S A* **85**(18): 6987-91.

211. Snapper, SB, RE Melton, S Mustafa, T Kieser and WR Jacobs, Jr. (1990) Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. *Mol Microbiol* **4**(11): 1911-9.

212. Springer, B, YG Kidan, T Prammananan, K Ellrott, EC Bottger and P Sander (2001) Mechanisms of streptomycin resistance: selection of mutations in the 16S rRNA gene conferring resistance. *Antimicrob Agents Chemother* **45**(10): 2877-84.

213. Springer, B, P Sander, L Sedlacek, WD Hardt, V Mizrahi, P Schar and EC Bottger (2004) Lack of mismatch correction facilitates genome evolution in mycobacteria. *Mol Microbiol* **53**(6): 1601-9.

214. Sreevatsan, S, KE Stockbauer, X Pan, BN Kreiswirth, SL Moghazeh, WR Jacobs, Jr., A Telenti and JM Musser (1997) Ethambutol resistance in Mycobacterium tuberculosis: critical role of embB mutations. *Antimicrob Agents Chemother* **41**(8): 1677-81.

215. Stahl, FW, MM Stahl and RE Malone (1978) Red-mediated recombination of phage λ in a recA- recB- host. *Mol Gen Genet* **159**: 207-211.

216. Stahl, MM, L Thomason, AR Poteete, T Tarkowski, A Kuzminov and FW Stahl (1997) Annealing vs. invasion in phage lambda recombination. *Genetics* **147**(3): 961-77.

217. Stover, CK, VF de la Cruz, TR Fuerst, JE Burlein, LA Benson, LT Bennett, GP Bansal, JF Young, MH Lee, GF Hatfull and et al. (1991) New use of BCG for recombinant vaccines. *Nature* **351**(6326): 456-60.

218. Susskind, MM and D Botstein (1978) Molecular genetics of bacteriophage P22. *Microbiol Rev* **42**(2): 385-413.

219. Swaminathan, S, HM Ellis, LS Waters, D Yu, EC Lee, DL Court and SK Sharan (2001) Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis* **29**(1): 14-21.

220. Takahashi, N and I Kobayashi (1990) Evidence for the double-strand break repair model of bacteriophage lambda recombination. *Proc Natl Acad Sci U S A* **87**(7): 2790-4.

221. Templin, A, SR Kushner and AJ Clark (1972) Genetic analysis of mutations indirectly suppressing recB and recC mutations. *Genetics* **72**(2): 105-15.

222. Thaler, DS, MM Stahl and FW Stahl (1987) Evidence that the normal route of replication-allowed Red-mediated recombination involves double-chain ends. *Embo J* **6**(10): 3171-6.

223. Thomason, L, DL Court, M Bubunenko, N Costantino, H Wilson, S Datta and A Oppenheim (2007) Recombineering: genetic engineering in bacteria using homologous recombination. *Curr Protoc Mol Biol* **Chapter 1**: Unit 1 16.

224. Thresher, RJ, AM Makhov, SD Hall, R Kolodner and JD Griffith (1995) Electron microscopic visualization of RecT protein and its complexes with DNA. *J Mol Biol* **254**(3): 364-71.

225. Tolun, G (2007) More than the sum of its parts: physical and mechanistic coupling in the phage lambda red recombinase. University of Miami. Ph.D. dissertation.

226. Turenne, CY, R Wallace, Jr. and MA Behr (2007) Mycobacterium avium in the postgenomic era. *Clin Microbiol Rev* **20**(2): 205-29.

227. van Kessel, JC and GF Hatfull (2007) Recombineering in Mycobacterium tuberculosis. *Nat Methods* **4**(2): 147-52.

228. van Kessel, JC and GF Hatfull (2008) Efficient point mutagenesis in mycobacteria using single-stranded DNA recombineering: characterization of antimycobacterial drug targets. *Mol Microbiol* **67**(5): 1094-107.

229. van Kessel, JC and GF Hatfull (2008) Mycobacterial recombineering. *Methods Mol Biol* **435**: 203-15.

230. Venkatesh, R, N Ganesh, N Guhan, MS Reddy, T Chandrasekhar and K Muniyappa (2002) RecX protein abrogates ATP hydrolysis and strand exchange promoted by RecA: insights into negative regulation of homologous recombination. *Proc Natl Acad Sci U S A* **99**(19): 12091-6.

231. Venkatesh, TV and CM Radding (1993) Ribosomal protein S1 and NusA protein complexed to recombination protein beta of phage lambda. *J Bacteriol* **175**(6): 1844-6.

232. Vilcheze, C, F Wang, M Arai, MH Hazbon, R Colangeli, L Kremer, TR Weisbrod, D Alland, JC Sacchettini and WR Jacobs, Jr. (2006) Transfer of a point mutation in Mycobacterium tuberculosis inhA resolves the target of isoniazid. *Nat Med* **12**(9): 1027-9.

233. Wagner, PL and MK Waldor (2002) Bacteriophage control of bacterial virulence. *Infect Immun* **70**(8): 3985-93.

234. Wang, J and KM Derbyshire (2004) Plasmid DNA transfer in Mycobacterium smegmatis involves novel DNA rearrangements in the recipient, which can be exploited for molecular genetic studies. *Mol Microbiol* **53**(4): 1233-41.

235. Wards, BJ and DM Collins (1996) Electroporation at elevated temperatures substantially improves transformation efficiency of slow-growing mycobacteria. *FEMS Microbiol Lett* **145**(1): 101-5.

236. Warming, S, N Costantino, DL Court, NA Jenkins and NG Copeland (2005) Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res* **33**(4): e36.

237. Weaver, S and M Levine (1977) Recombinational circularization of Salmonella phage P22 DNA. *Virology* **76**(1): 29-38.

238. Weaver, S and M Levine (1977) The timing of erf-mediated recombination in replication, lysogenization, and the formation of recombinant progeny by Salmonella phage P22. *Virology* **76**(1): 19-28.

239. Wong, I and TM Lohman (1993) A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proc Natl Acad Sci U S A* **90**(12): 5428-32.

240. Yu, D, HM Ellis, EC Lee, NA Jenkins, NG Copeland and DL Court (2000) An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A* **97**(11): 5978-83.

241. Yu, D, JA Sawitzke, H Ellis and DL Court (2003) Recombineering with overlapping single-stranded DNA oligonucleotides: testing a recombination intermediate. *Proc Natl Acad Sci U S A* **100**(12): 7207-12.

242. Zhang, Y, F Buchholz, JP Muyrers and AF Stewart (1998) A new logic for DNA engineering using recombination in Escherichia coli. *Nat Genet* **20**(2): 123-8.

243. Zhang, Y, JP Muyrers, G Testa and AF Stewart (2000) DNA cloning by homologous recombination in Escherichia coli. *Nat Biotechnol* **18**(12): 1314-7.

244. Zhang, Y, JP Muyrers, J Rientjes and AF Stewart (2003) Phage annealing proteins promote oligonucleotide-directed mutagenesis in Escherichia coli and mouse ES cells. *BMC Mol Biol* **4**(1): 1.

245. Zuniga-Castillo, J, D Romero and JM Martinez-Salazar (2004) The recombination genes addAB are not restricted to gram-positive bacteria: genetic analysis of the recombination initiation enzymes RecF and AddAB in Rhizobium etli. *J Bacteriol* **186**(23): 7905-13.